

데이터 마이닝 기법을 이용한 항만물동량 예측 활용방안 연구

2006. 12

전찬영 · 송주미

□ 보고서 집필 내역

◆ 연구책임자

- 전 찬 영 : 제1장, 제3장~제5장
제7장, 총괄

◆ 연구진

- 송 주 미 : 제6장 일부

◆ 협동 연구진

- 김 현 중 (연세대학교 교수) : 제2장, 제6장

□ 산·학·연·정 연구자문위원

◆ 김 창 곤 (순천대학교 교수)

◆ 신 승 식 (여수대학교 교수)

◆ 남 재 현 (해양수산부 항만정책과 사무관)

* 연구자문위원은 산·학·연·정 순임

머 리 말

우리나라의 항만물동량은 개발연대인 1970~1980년대를 지나 '90년대 이후에도 연평균 두 자릿수의 높은 급증세를 이어갔다. 2000년대 초반까지만 해도 연 10%대 이상의 증가세를 유지함으로써 물동량의 성장세는 지속되는 듯 보였으나 이후 증가세가 급격히 둔화되는 조짐을 보이고 있다. 이러한 증가세 둔화의 원인으로는 많은 요인이 거론되고 있다. 대내적으로는 그동안 물동량의 대량 창출원이었던 전통 제조업의 성장세 둔화와 첨단 신산업의 부상, 서비스 산업의 비중 확대 등으로 대표되는 산업구조의 변화가 가장 큰 요인으로 꼽히고 있다. 대외적으로는 수출입의 대외적 환경 악화, 상해 양산 항을 비롯한 중국의 대규모 항만 개발, 중국의 다국적 물류기지화 등이 지적되고 있다.

이와 같이 국내에서 처리되는 항만물동량의 총체적 변화가 예감되고 있으나 5년 단위로 계획의 수립 내지 보완·수정을 거치도록 되어 있는 정부의 항만개발기본계획에는 별다른 대응방안이 나오지 않고 있다. 정책의 일관성 유지를 위해 한 번 수립된 계획은 단기간의 물동량 변화만을 토대로 전면 손질할 수 없다는 정부 나름의 입장도 충분히 이해할 수 있다. 이러한 정부의 경직된 자세는 기본계획의 수립 혹은 재점검 간의 기간 간격이 길다는 것과 항만물동량에 대한 예측의 초점이 너무 중장기적인 것에 맞춰져 있어 계획 조정에 대한 융통성이 부족한 점에 기인한다.

우리나라 항만개발의 근거가 되는 항만개발기본계획은 항만법상 10년을 주기로 수립되며 재점검은 5년마다 이루어지고 있다. 물동량 예측의 목표 연도는 예측시점부터 10년 이상 20년까지에 맞춰져 있다. 따라서 지금처럼 대내외적 물류환경이 급변하고 있는 상황에서 먼 장래의 물동량을 정확히 예측한다는 것은 기대하기 어렵고, 5년 단위의 재점검 기간도 변화되는 여건을 반영하기에는 그 간격이 너무 길다.

해양수산부가 당 원에 ‘수요예측센터’를 설치하여 항만물동량의 변동성 감지 및 이를 실시간 예측에 반영할 수 있도록 한 것도 정부의 장기 항만개발계획을 좀 더 탄력적으로 운영하자는 뜻이 담겨 있다. 항만물동량에 영향을 미칠 수 있는 대내외적 요인의 변동성을 지속적으로 반영하고, 나아가 장기적인 항만개발계획까지 융통성 있게 운용할 수 있기 위해서는 물동량에 대한 단기적인 예측이 정기적으로 이루어질 필요가 있다. 월별 혹은 분기별 자료를 이용한 단기적인 예측은 큰 데이터의 용량과 더욱 정밀한 정확성을 요구하게 된다. 이런 점에서 데이터의 일정한 추세를 전제, 중장기적인 시점에서 예측해 왔던 기존의 예측방법(주로 회귀모형 혹은 변형회귀모형)과는 구분될 수 있는 보다 정교한 예측기법의 사용 필요성이 제기되고 있다.

최근 수리 예측방법으로 데이터 마이닝(Data Mining)에 의한 예측기법이 널리 활용되고 있다. 데이터 마이닝은 일반적으로 데이터베이스에 있는 방대한 양의 정보에 숨어 있는 지식을 자동적으로 추출하는 과정이다. 데이터 마이닝의 목적은 사용자에게 유용한 새로운 지식을 생성하는 것으로 수립된 모형에 따라 크게 서술모형과 예측모형으로 나눈다. 인공신경망모형(Neural Network Model)은 예측모형의 대표적인 기법으로 과거에 수집된 데이터로부터 반복적인 학습과정을 거쳐 데이터에 내재되어 있는 패턴을 찾아내는 모델링 기법이다. 이 기법은 예측력이 우수하다는 장점을 가지고 있으나 적용한 함수모형을 나타내지 않아 결과에 대한 이유를 설명하지 못하는 단점을 가지고 있다.

본 연구는 최근 다양한 분야에서 주목받고 있는 데이터 마이닝 기법에 의한 수리 예측방법 중 가장 널리 이용되고 있는 신경망분석모형의 활용에 초점을 맞추고 있다. 즉, 이러한 수리예측방법을 항만물동량에 적용시켜 이를 예측해 보고, 예측의 신뢰도 및 정확도를 비롯한 다양한 통계적 특성들을 기존 모형과 비교해 봄으로써 동 기법의 유용성을 검증하고자 하였다. 궁극적으로는 기존에 사용되어 왔던 회귀모형과 시계열모형의 신뢰도 향상을 위한 개선 대안으로 데이터 마이닝의 대표적 예측기법인 신경망분석모형의 활용방안을 강구하는 데에 있다.

본 연구보고서는 본 원의 전찬영 연구위원이 총괄 집필하였고, 송주미 연구원이 공동 집필하였다. 올 여름 연구원을 퇴사한 정혜원 전 연구원은 재직시 선행연구를 비롯한 많은 자료를 수집하여 본 연구의 진행에 큰 도움을 주었다. 그리고 연세대학교의 김현중 교수가 협동연구진으로 참여하여 모형의 신뢰도와 강건성을 비교하기 위한 몬테카를로 시뮬레이션 분석을 하였다. 본 연구의 수행과정에서 자문을 해주신 전남대학교의 신승식 교수, 순천대학교의 김창곤 교수, 본 원의 정봉민 선임연구위원, 그리고 양창호 기획조정실장님께 깊은 사의를 표한다.

2006년 12월

韓國海洋水產開發院
院長 李 正 煥

목 차

ABSTRACT	i
----------------	---

요 약	v
-----------	---

제1장 서 론	1
---------------	---

1. 연구의 필요성 및 목적	1
1) 연구의 필요성	1
2) 연구의 목적	2
2. 연구 범위와 방법	4

제2장 데이터 마이닝 기법에 대한 정의	7
-----------------------------	---

1. 데이터 마이닝의 개념	7
1) 데이터 마이닝의 기능	8
2) 데이터 마이닝 응용분야	9
2. 데이터 마이닝 분석 절차	11
1) 샘플링/선택	11
2) 탐색	11
3) 변형과 조정	12
4) 모형화	12
5) 평가	13

3. 데이터 마이닝 기법 소개	13
1) 기술모형	13
2) 예측모형	14
4. 인공지능망의 특성과 차이점	20
1) 인공지능망의 특성	20
2) 인공지능망과 기타 예측방법과의 차이점	21
제3장 국내외 선행연구 고찰	23
1. 회귀모형과의 비교 연구	23
1) 연구의 주요내용 및 특징	23
2) 시사점	25
2. 시계열모형과의 비교 연구	26
1) 연구의 주요내용 및 특징	26
2) 시사점	28
3. 기타 관련 연구	28
1) 시사점	29
4. 선행연구와의 차이점	30
제4장 주요 물동량 추세 변화 분석 및 설명변수 설정	32
1. 주요 품목의 선정	32
2. 품목별 물동량의 추세적 변화 특징 분석 및 설명변수 설정	36
1) 품목별 항만물동량 추세 변화의 구분	36
2) 수입물동량	38
3) 수출물동량	55
제5장 품목별 예측의 적합도 비교분석	73

1. 방법론	73
2. 적합도 비교·분석	76
1) Regression과 N.N의 결과 비교	76
2) 변형 Regression과 N.N의 결과 비교	86
3) 예측 결과 비교 종합	90
 제6장 몬테카를로 시뮬레이션 분석	95
1 방법론	95
2. 시뮬레이션 설계	95
3. 신뢰성 비교	98
4. 강건성 비교	104
5. 비교 종합	111
 제7장 결론 및 정책대안	112
1. 결론	112
2. 정책대안	114
 참고문헌	117
 부 록	119

표목차

〈표 2-1〉 데이터 마이닝의 응용분야	19
〈표 3-1〉 실제값에 예측값의 비율(161cm 기준)	25
〈표 3-2〉 모형간 예측력 비교	27
〈표 4-1〉 품목별 수입금액 비율 변동추이	34
〈표 4-2〉 품목별 수출금액 비율 변동추이	35
〈표 5-1〉 품목별 물동량과 설명변수	75
〈표 5-2〉 예측의 적합도 비교분석 I (Regression과 Neural Network)	78
〈표 5-3〉 변형회귀모형	88
〈표 5-4〉 예측의 적합도 비교분석 II(변형 Regression과 Neural Network)	89
〈표 6-1〉 모형별 신뢰성 비교분석(Regression과 Neural Network)	103
〈표 6-2〉 모형별 강건성 비교분석(Regression과 Neural Network)	107

그림목차

〈그림 1-1〉 연구 추진 흐름도	6
〈그림 2-1〉 데이터 마이닝 모형의 종류	9
〈그림 2-2〉 인공신경망모형의 구성	17
〈그림 2-3〉 시그모이드 함수의 형태	18
〈그림 2-4〉 인공신경망모형화의 과정	18
〈그림 4-1〉 전기기기 수입물동량 변동추이	39
〈그림 4-2〉 전기기기 수입물동량 및 관련 변수 변동추이	39
〈그림 4-3〉 기계류 수입물동량 변동추이	40
〈그림 4-4〉 기계류 수입물동량 및 관련 변수 변동추이	41
〈그림 4-5〉 화학공업생산물 수입물동량 변동추이	42
〈그림 4-6〉 화학공업생산물 수입물동량 및 관련 변수 변동추이	43
〈그림 4-7〉 철강 수입물동량 변동추이	44
〈그림 4-8〉 철강 수입물동량 및 관련 변수 변동추이	45
〈그림 4-9〉 비철금속 수입물동량 변동추이	46
〈그림 4-10〉 비철금속 수입물동량 및 관련 변수 변동추이	47
〈그림 4-11〉 기타 수입물동량 변동추이	48
〈그림 4-12〉 기타 수입 물동량 및 관련 변수 변동추이	48
〈그림 4-13〉 방직용섬유 수입물동량 변동추이	49
〈그림 4-14〉 방직용섬유 수입물동량 및 관련 변수 변동추이	50
〈그림 4-15〉 플라스틱고무 수입물동량 변동추이	51
〈그림 4-16〉 플라스틱고무 수입물동량 및 관련 변수 변동추이	52
〈그림 4-17〉 기타광석 수입물동량 변동추이	53
〈그림 4-18〉 기타광석 수입물동량 및 관련 변수 변동추이	53

〈그림 4-19〉 유연탄 수입물동량 변동추이	54
〈그림 4-20〉 유연탄 수입물동량 및 관련 변수 변동추이	55
〈그림 4-21〉 전기기기 수출물동량 변동추이	56
〈그림 4-22〉 전기기기 수출물동량 및 관련 변수 변동추이	57
〈그림 4-23〉 기계류 수출물동량 변동추이	58
〈그림 4-24〉 기계류 수출물동량 및 관련 변수 변동추이	59
〈그림 4-25〉 자동차 수출물동량 변동추이	60
〈그림 4-26〉 자동차 수출물동량 및 관련 변수 변동추이	61
〈그림 4-27〉 플라스틱고무 수출물동량 변동추이	62
〈그림 4-28〉 플라스틱고무 수출물동량 및 관련 변수 변동추이	62
〈그림 4-29〉 철강 수출물동량 변동추이	64
〈그림 4-30〉 철강 수출물동량 및 관련 변수 변동추이	64
〈그림 4-31〉 화학공업생산품 수출물동량 변동추이	66
〈그림 4-32〉 화학공업 수출물동량 및 관련 변수 변동추이	66
〈그림 4-33〉 방직용섬유 수출물동량 변동추이	67
〈그림 4-34〉 방직용섬유 수출물동량 및 관련 변수 변동추이	68
〈그림 4-35〉 기타 수출물동량 변동추이	69
〈그림 4-36〉 기타 수출물동량 및 관련 변수 변동추이	69
〈그림 4-37〉 비철금속 수출물동량 변동추이	70
〈그림 4-38〉 비철금속 수출물동량 및 관련 변수 변동추이	71
〈그림 4-39〉 조제식품 수출물동량 변동추이	72
〈그림 4-40〉 조제식품 수출물동량 및 관련 변수 변동추이	72
〈그림 5-1〉 방직용섬유 및 그 제품(수입)의 실적치 및 예측치 비교	81
〈그림 5-2〉 비철금속 및 그 제품(수입)의 실적치 및 예측치 비교	81
〈그림 5-3〉 전기기기 및 그 제품(수입)의 실적치 및 예측치 비교	82
〈그림 5-4〉 기타(수입)의 실적치 및 예측치 비교	82
〈그림 5-5〉 유연탄(수입)의 실적치 및 예측치	84

〈그림 5-6〉 전기기기 및 그 부품(수출)의 실적치 및 예측치	85
〈그림 5-7〉 플라스틱고무 및 그 부품(수출)의 실적치 및 예측치	85
〈그림 5-8〉 ‘상승 지속형’ 품목의 2001~2004 실제치와 예측치의 비교	93
〈그림 5-9〉 ‘불규칙 상승형’ 품목의 2001~2004 실제치와 예측치의 비교	93
〈그림 5-10〉 ‘상승 둔화형’ 품목의 2001~2004 실제치와 예측치의 비교	94
〈그림 5-11〉 ‘정체 혹은 불규칙형’ 품목의 2001~2004 실제치와 예측치의 비교	94
〈그림 6-1〉 수입_기타와 수입_유연탄 품목의 회귀분석모형 예측값과 95% 몬테카를로 신뢰구간	97
〈그림 6-2〉 수입_플라스틱 고무와 수출_조제식품 품목의 신경망분석 모형 예측값과 95% 몬테카를로 신뢰구간	98
〈그림 6-3〉 수입_기계류 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 비교	99
〈그림 6-4〉 수입_기계류 품목의 회귀분석모형과 신경망분석모형모형의 표준편 차 비교	100
〈그림 6-5〉 수출_기계류 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 비교	100
〈그림 6-6〉 수출_기계류 품목의 회귀분석모형과 신경망분석모형의 표준편차 비교	101
〈그림 6-7〉 수출_조제식품 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 비교	101
〈그림 6-8〉 수출_조제식품 품목의 회귀분석모형과 신경망분석모형의 표준편차 비교	102
〈그림 6-9〉 수입_철강 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 비 교	108

〈그림 6-10〉 수입_철강 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 길이 비교	108
〈그림 6-11〉 수출_기계 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 비교	109
〈그림 6-12〉 수출_기계 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 길이 비교	109
〈그림 6-13〉 수출_철강 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 비교	110
〈그림 6-14〉 수출_철강 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 길이 비교	110

ABSTRACT

Analysis of Port Cargo through Use of Data Mining Techniques

In the 1990s, port cargoes in Korea were increased by double digit percentages every year, and in the earlier period of the 2000s the rapid growth of port cargo continued. But thereafter, its growth seemed to drop suddenly due to various factors. These factors could be examined from two points of view. From an internal viewpoint, the structure of industry changed drastically. It was represented by such trends as the weakening of traditional industry, which was the source of mass seaborne cargoes, the prominence of new industry with advanced technology, and the enlargement of the proportion of the service industry in the economy. From a foreign perspective, the worsening of the environment for foreign trade, the mass construction of the port facilities and the expansion of the international logistics base in China were evident.

In Korea, a national master plan, called the “Basic development plan for national major ports” is established every 10 years according to port legislation. Also, it should be reevaluated every five years. Therefore, the forecast for port cargo transit cargo is on a long-term basis, over 10~20 years. The long-term period between founding of the master plan and reevaluation causes the forecast to be inaccurate because it is too long to

consider all the changes of extraneous factors in such rapidly changing circumstances.

To resolve this problem, the Ministry of Marine and Fisheries (MOMAF) founded the 'Forecasting Center for Demand' by establishing the Korea Maritime Institute (KMI). The role of this center is to monitor all the variables related to port cargoes and reflect them in its short-term forecast. Consequently, it contributes to the flexible execution of the long-term port construction plan.

The forecast, focused on the short-term, should be based on the recent raw data, such as monthly or quarterly observations and needs to be more accurate than forecasted, focused on the long term. But, to make it more precise, a more accurate form of forecasting model is needed. Among existing models of forecasting, the regression or time series model is most widely used. It lays stress on the historic trends of available data. Its weakness is that the future changes may be different from historic trends of given data or it might even move in unexpected directions. In addition, forecasting port cargoes based on monthly or quarterly data, the number of available data being increased, we should use visual models for this type of forecasting.

In recent times, the forecasting of models using data mining has become more noticeable. Data mining is generally treated as being a process for inducing knowledge automatically from a magnificent database. It applies to the description of certain states or forecasts. The Neural Network Model (NNM) is a representative forecasting model utilizing the Data Mining technique. It finds certain kind of patterns from given data by repeating learning processes. In forecasts, this model tends to be more precise than any other. But it doesn't show any function used for forecasting, so it

cannot explain the relationship (cause and effect) between variables.

In this thesis, we were interested in NNM as a representative forecasting model using Data Mining techniques and focused on its practical application. We forecasted port cargo throughput using NNM and compared it with other results from the existing model (regression model and transformed regression model). In comparing forecast models, the forecasted port cargo items were categorized into four types according to its moving pattern of given data. The four types are: 'growth continuing', 'growth becoming dull', 'move irregularly but growing in stepwise', and 'static but moving irregularly'. Statistical criteria judging the priority between models are: fitness, reliability and robustness. For the simulation model, the Monte Carlo method was used.

In this result, the regression model was more suitable to the items where trends were growing continuously, regardless of its growth pattern. On the other hand, if the moving trends fell into the categories of 'growth becoming dull' or 'static but moving irregularly', NNM was rated superior to other models. In a rapidly changing era, we could expect that NNM would be widely applied for forecasting models. However, it could be used only in case in which accurate forecast could be more important due to its inability to interpret the relationships among variables.

From this study, NNM could be more effectively used in short-term forecasts, because considering various reasons, the preciseness of forecasting was more important than the grasp of its causality. As supplementary tools to discover the relationships among variables, the □□decision-tree□□ would be used.

But, in forecasting focused on the long-term, things are different. Forecast of port cargo throughput in the long-term context are closely related to the

national development plan and the policy direction of the government. Therefore, concrete steps to reflect the changes in the economic factors, which explain the aforementioned policy factors, tends to be more important. In this thesis, as an alternative to apply NNM for long-term forecasting, NNM and regression models were suggested to be used in a compositive way.

제1장 서론

- 항만개발계획의 기본이 되는 물동량 수요 전망을 위해서는, 물동량의 변동 요인을 정확히 파악하여 예측모형을 설정하는 것이 요구됨
 - 항만개발의 규모가 매년 수천억 원대에 이르며 개발에 소요되는 기간이 길기 때문임
 - 물동량의 수요를 정확히 예측하기 위해서는 물동량의 추세적 변화와 이에 영향을 미치는 대내외적 요인을 정확히 파악하고 이를 반영해 줄 수 있는 예측모형을 설정하는 것이 중요함
- 물동량에 영향을 미치는 경제·사회적 여건의 최근의 급속한 변화를 제대로 반영할 수 있는 다양한 예측방법론을 개발할 필요가 있음
 - '90년대 이후 급증세를 보여 왔던 우리나라 컨테이너물동량은 2000년대 초반 이후 증가세가 급격히 둔화되는 조짐을 보이고 있음
 - 과거 데이터의 추세적 변화요인 외에, 최근 나타나고 있는 기술혁신 및 산업구조조정에 따른 물동량 특성의 변화를 반영할 수 있어야 함
 - 또한 중장기적인 예측뿐 아니라, 월별 혹은 분기별 자료를 이용한 1~2년 단위의 단기적인 예측이 행해질 필요가 있음
- 본 연구는 데이터 마이닝(Data Mining) 기법 중 인공신경망모형(Neural Network Model, NNM)에 초점을 맞추었음
 - 데이터 마이닝 기법은 최근 다양한 분야에서 활용도를 높이고 있으며, 그 중 가장 많이 이용되는 기법은 인공신경망모형임
 - 인공신경망모형을 통하여 항만물동량을 예측해보고, 예측의 신뢰도 및 정확도 등을 기존 예측모형과 비교·평가하였음

- 기존 모형과 인공신경망모형을 품목별로 비교하였으며, 품목별 물동량의 변화유형을 크게 4가지 유형으로 분류
 - 통상적으로 항만물동량의 예측은 품목별로 이루어지므로, 기존모형과 인공신경망모형을 품목별로 예측의 적합도(fitness), 신뢰성(reliability), 강건성(robustness)을 기준으로 비교
 - 또한 품목별 물동량의 변화유형을 4가지 범주로 분류하여, 인공신경망모형의 적절한 예측 활용방안을 모색

제2장 데이터 마이닝 기법에 대한 정의

- 데이터 마이닝은 일반적으로 데이터베이스에 있는 방대한 양의 정보에 숨어있는 지식을 자동적으로 추출하는 과정임
 - 데이터에 숨겨진 패턴과 관계를 찾아내어 광맥을 찾아내듯이 정보를 발견해 내는 것임
 - 여기서 정보 발견이란 데이터에 고급 통계 분석과 모델링 기법을 적용하여 유용한 패턴과 관계를 찾아내는 과정임
- 데이터 마이닝은 현재 특정분야에서 일반화되어 있지만, 점점 다양한 분야로 확장되고 있는 추세임
 - 현재 데이터베이스 마케팅의 핵심 기술이 되어가고 있으며, 소매업, 금융업, 물류/교통 등 다른 분야에도 응용이 되고 있으나 아직은 초기 단계에 머물러 있는 상태임
 - 그러나 데이터 웨어하우스 시스템 구축 및 소프트웨어 출시 등으로 미루어볼 때, 곧 일반화될 것으로 판단됨
- 데이터 마이닝의 여러 기법 중 대표적인 것으로는 의사결정나무와 인공신경망모형을 들 수 있으며, 각각 장단점이 있음
 - 의사결정나무는 이해하기 쉽다는 장점이 있으나, 연속형 데이터를 처리하는 능력이 떨어진다는 단점이 있음

- 인공신경망모형은 입력변수와 목적변수의 관계를 그리기가 어려운 복잡한 데이터에 대해서도 좋은 예측력을 보여 주는 방법임
- 그러나 결과물만을 제공할 뿐, 어떤 변수가 중요한지 어떻게 상호작용이 이루어지는지에 대한 설명을 하지 않는다는 단점이 있음
- 즉 인공신경망모형은 중간과정은 알 수 없지만, 변수의 형태에 구애받지 않으며 예측력이 우수한 방법임
- 인공신경망모형은 과거에 수집된 데이터로부터 반복적인 학습 과정을 거쳐 패턴을 찾아내고 이를 일반화하는 방법임
- 인공신경망모형은 인간 두뇌의 신경세포를 모방하여 마디(node)와 고리(link)로 구성된 망구조를 모형화하는 방법임
- 입력층, 은닉층, 출력층으로 구성되어 있으며, 각 층은 노드로 이루어져있고, 서로 연결되어 있음
- 입력층을 통해 변수를 입력하면, 은닉층에서 변수 값들의 선형결합을 비선형 결합으로 처리하여, 출력층이나 다른 은닉층에 전달하며, 이러한 과정을 거쳐 최종적으로 출력층을 통하여 예측값이 생성됨
- 은닉층에서 비선형 결합을 처리할 때 이용되는 함수를 활성화함수라 하는데, 흔히 시그모이드(sigmoid) 함수가 많이 사용함
- 비선형함수에 의한 예측값은 선형함수에 의한 예측값보다 더 정교하다고 알려져 있듯, 예측력은 인공신경망모형의 최대 장점임

제3장 국내외 선행연구 고찰 및 차이점

- 다양한 분야에서 회귀분석 및 시계열 분석 등 기존의 통계모형과 인공신경망모형의 우수성을 비교하고 있으며, 주로 단기간 내에 심한 변동을 보이는 자료에서 인공신경망모형의 예측력이 두드러짐
- Lam, W. H. K. and Pan L. P. and Eddie C. M(2004)은 홍콩의 품목별 물동량을 회귀분석과 인공신경망모형으로 각각 예측하여, 적합도 및 신뢰성을 비교·분석

- 1983~2000년 동안의 18개 시계열 자료를 이용하여 2002~2011년의 물동량을 예측
- 적합도 검증을 위해서는 2001년의 실측치를 이용하여 회귀분석 결과와 인공신경망모형의 결과 비교
- 아파트와 홍수위 예측분야에서 인공신경망모형과 회귀모형을 비교한 결과 인공신경망모형의 정확도가 더 높게 나타남
 - 김태훈·홍한국(2004)은 아파트 가격 예측 분야에서 인공신경망모형과 회귀모형의 정확도를 비교
 - 손민우·이길성(2003)은 홍수위 예측 분야에서 신경망 이론과 다중회귀분석 방법의 적합성을 비교
- 노태협·이택호·한인구(2005)는 KOSPI1200 주가지수의 변동예측에 시계열의 GARCH 모형과 인공신경망모형을 비교
 - 변동 방향의 예측에서는 GARCH가, 변동 예측의 정확성에서는 인공신경망이 우수하게 나타남
- 정효준·이홍근(2002)은 BOD 예측에 시계열의 ARIMA 모형과 인공신경망모형의 적정성을 비교
 - 평균을 중심으로 변동하는 기간에는 ARIMA 모형이, 최대와 최소값을 나타내는 기간에는 인공신경망모형이 보다 잘 부합함
- 선행연구들은 통계적 예측방법론과 인공신경망모형, 두 가지 방법을 상호보완적으로 사용할 것을 제안
 - 인공신경망모형을 이용한 예측은 정확도 측면에서 우수한 경향이 있지만, 과정에 대한 설명이 부족하다는 단점이 있으므로
 - 기존의 통계모형과 인공신경망모형을 혼합하여 이용하면, 변수 간의 상호 영향력 분석 및 예측력 향상을 동시에 추구할 수 있기 때문
- 본 연구는 회귀분석과 인공신경망모형을 비교하되 신뢰성(Reliability)과 강건성(Robustness)까지 고려하였으며, 유형별 분석을 시도하였음

- 회귀분석과 인공신경망모형의 적합도 비교라는 측면에서만 본다면 본 연구는 선행연구들과 유사함
- 그러나 대부분의 연구가 예측의 적합도만을 비교의 잣대로 할 뿐, 신뢰성(Reliability)을 고려하는 경우는 있어도 예측의 민감도를 측정하는 강건성(Robustness)에 대한 고려는 하지 않고 있음
- 본 연구와 가장 유사한 Lam, W. H. K. and Pan L. P. and Eddie C. M(2004)의 연구는, 품목별로 어떠한 변화 특성을 가진 물동량의 예측에 적합한지에 대한 분석을 수행하지 않았음
- 또한 본 연구는 변형회귀모형도 인공신경망모형과 비교하였음
 - 인공신경망모형과의 비교대상을 단순히 회귀모형에 국한하지 않고, 실제 예측에 많이 활용되고 있는 변형회귀모형도 포함
 - 이를 통해 변수 변환을 통한 변형회귀모형이 단순 회귀모형에 비해 어떠한 변화를 가져왔는지를 분석

제4장 주요 물동량 추세 변화 분석 및 설명변수 설정

- 실증분석에 사용될 주요 품목으로 수입과 수출에서 각각 10개씩 선정
 - 「해양수산통계연보」 상에 분류된 32개의 품목을 대상으로 하여, 2005년 기준으로 비중이 높은 순서로 상위 10개씩을 선정
 - 품목별 금액산출은 관세청의 HS 코드를 「해양수산통계연보」 상의 32개 품목 분류에 대응시켜 구하였음
 - 단, 하역과정에서 안벽시설이 필요 없는 ‘유류’와 물동량 규모가 작아 항만시설 소요가 크게 필요하지 않은 ‘항공기/선박 및 그 부품’은 품목 선정에서 제외
- 수입 10개 품목은 전체 수입금액의 65.6%, 수출 10개 품목은 각각 전체 수출금액의 87.1%를 차지함
 - 항만시설소요 산정에서 제외되는 석유류 품목을 제외하면 10개 품목이 수입과 수출에서 차지하는 비중은 각각 78.4%와 92.1%로 높아짐

- 물동량으로 환산할 경우 이들 10개씩의 품목이 석유류를 제외한 전체 물동량에서 차지하는 비중은 수입과 수출이 각각 60.3%와 91.1%임
- 품목별 물동량의 변화추세는 다음의 4가지 변동유형을 설정
 - 상승 지속형 : 물동량 추세가 전반적으로 지속적인 상승을 하는 형태
 - 상승 둔화형 : 전반적으로 상승하나 증가율이 차츰 낮아지는 형태
 - 불규칙 상승형 : 구간(단기)별로 변동이 불규칙적이거나 단계별 저점이 높아지는 형태
 - 정체 또는 불규칙 변동형 : 정체 내지 평균값을 중심으로 상하 변동을 하나 전반적으로 증가나 감소를 하지 않는 형태
- 상승 지속형은 수입 5개, 수출 2개 품목이 각각 설정되었음
 - 수입 : '기계류 및 그 부품', '철강 및 그 제품', '방직용 섬유 및 그 제품', '기타광석 및 생산물', '유연탄'
 - 수출 : 차량 및 그 부품, '방직용 섬유 및 그 제품'
- 상승 둔화형은 수입 2개, 수출 5개 품목이 각각 설정되었음
 - 수입 : 전기기기 및 그 부품, '화학공업생산물'
 - 수출 : '플라스틱/고무 및 그 제품', '철강 및 그 제품', '화학공업생산물', '비철금속 및 그 제품', '조제식품/음료/주류'
- 불규칙 상승형은 수입 1개, 수출 2개 품목이 각각 설정되었음
 - 수입 : '비철금속 및 그 제품'
 - 수출 : '전기기기 및 그 부품', '기계류 및 그 부품'
- 정체 또는 불규칙 변동형은 수입 2개, 수출 1개 품목의 각각 설정되었음
 - 수입 : '기타', '플라스틱/고무 및 그 제품'
 - 수출 : '기타'
- 추후 회귀분석방법과 인공신경망모형의 예측 결과 비교시 이러한 변동 유형과 연관시켜 분석
 - 이는 회귀분석과 인공신경망모형의 예측력이 물동량의 변화추세 형태

에 크게 영향을 받을 것이라는 가정이 전제된 것임

- 본 원에서 기 발간한 연구보고서의 내용을 토대로 당시 품목별 물동량 예측을 위해 사용하였던 설명변수(explanatory variable)와 품목별 물동량과의 관계를 재검토
 - 선행연구를 통하여 다수의 변수를 설명변수의 후보군으로 설정한 뒤
 - 그 중에서 예측변수와 설명변수의 통계적 유의성을 고려하여 최종 설명변수를 선택

제5장 품목별 예측의 적합도 비교·분석

- 기본적으로 회귀분석과 인공신경망모형, 추가로 본 원이 기존 연구보고서에서 사용하였던 변형된 회귀모형과 인공신경망모형의 적합도 비교·분석 실시
 - 예측의 적합도란 모형에 의한 예측 결과가 해당 연도의 실적치에 얼마나 잘 부합되는가를 의미
 - 변형된 회귀모형이란 종속변수 혹은 설명변수를 변수 변환하거나, 선형함수의 형태를 비선형 형태로 치환한 것을 의미
- 적합도 비교는 기본적으로 품목별로 이루어지나 모형의 유용성 평가를 위해 4가지 물동량의 변화유형을 분석의 기본으로 삼음
 - 적합도 비교는 각 품목별로 1981~2000년까지의 관찰개수를 사용하여 각각의 모형을 추정한 뒤
 - 2001~2004년의 4개 년도에 대해 추정모형에 의한 예측치와 실적치를 대비시켜 모형별 적합도를 비교·판정하였음
- 기준이 되는 통계적 특성치는 예측치와 실적치 차이의 평균값인 MAE(Mean Absolute Error)로 설정하였고, R^2 (결정계수)를 참고지표로 이용
- 각 품목별 물동량 예측에 사용된 설명변수는 기 발간된 물동량 예측보

고서의 것을 인용

- 회귀모형과 인공신경망모형 간의 적합도를 비교한 결과, 추세유형별로 확연히 구분되었음
- ‘상승 지속형’은 회귀모형이 인공신경망모형보다 우수하게 나타남
 - 7개 품목 중 수입 방직섬유를 제외한 6개 품목의 예측 적합도에서 회귀모형이 인공신경망모형보다 우수한 것으로 나타남
 - ‘상승 지속형’에 속하는 7개 품목을 합산할 경우에도 회귀모형의 MAE가 더 적어, 인공신경망모형보다 더 우수한 것으로 나타남
- ‘상승 둔화형’은 인공신경망모형이 회귀모형보다 적합도가 더 높은 것으로 나타남
 - 수입 전기기기와 수출 철강제품을 제외한 5개 품목에서 인공신경망모형이 회귀모형보다 더 적합도가 높은 것으로 나타남
 - ‘상승 둔화형’에 해당되는 모든 품목의 MAE를 단순 합산할 경우, 인공신경망모형이 더 적게 나타났기 때문에, 인공신경망모형을 사용하는 것이 예측의 적합도를 높일 수 있는 방안이 될 수 있음
- ‘불규칙 상승형’에 해당되는 3개 품목 중 2개 품목에서 회귀모형이 상대적으로 양호한 것으로 평가됨
 - 수출 기계류와 수출 전기기기는 회귀모형이, 수입 비철금속은 인공신경망모형이 더 양호한 것으로 나타남
 - ‘불규칙 상승형’의 품목 전체를 고려할 경우 회귀모형의 MAE가 더 적게 나타나 동 유형의 전반적인 예측모형은 회귀모형이 더 나은 것으로 파악됨
- ‘정체 또는 불규칙 변동형’은 객관적인 판단을 내리기 어려우나, 전체적으로 볼 때 비선형적인 관계를 고려할 수 있는 인공신경망모형이 더 적합하다고 할 수 있음
 - 산업의 구조가 급격히 변화되고 있는 시대 흐름에 비추어, 물동량의

- 추세를 가늠하기 어려운 품목의 수는 차츰 늘어날 전망이다
- 동 유형에 속한 품목이 3개밖에 되지 않아 품목별로 적합도에 대해 객관성 있는 판정을 내린다는 것이 별로 유의하지 않다고 판단됨
- 동 유형에 속한 3개 품목의 통계적 특성치를 비교할 경우, 인공신경망모형이 보다 우수한 것으로 나타남
- 각 유형별로 예외적인 사례가 존재하지만, 이런 경우에도 해당 품목의 물동량 추세유형에서 많은 공통적인 사유를 발견할 수 있었고, 회귀모형과 신경망모형의 적용 가능한 범위에 대한 시사 가능한 여러 요인들로 파악되었음
- 또한 변형된 회귀모형과 인공신경망모형의 비교·분석을 수행
 - 기존에 시도된 회귀모형의 변형 형태의 범위가 너무 넓어, 본 연구에서는 2004년 본 원의 용역연구보고서인 「전국 항만물동량 예측」상의 품목별 회귀예측모형을 그대로 적용
 - 변형된 회귀모형에 의한 예측치의 MAE는 단순 회귀모형의 MAE에 비해 전반적으로 낮은 상태를 유지하고 있으며, 신경망모형과의 차이도 감소한 것으로 나타남
- 변형된 회귀모형의 추세유형별 적합도의 우열관계는 전반적으로 크게 다를 바 없으나, 인공신경망모형에서 회귀모형으로 우열관계가 바뀌는 경우가 있음
 - ‘상승 둔화형’에 속하는 수출 플라스틱·고무 품목은 적합모형이 인공신경망모형에서 회귀모형으로 우열관계가 바뀌었음
 - ‘정체 또는 불규칙 변동형’에서는 수출기타잡화 품목을 제외한 나머지 두 품목에서 적합모형이 달라진 것으로 나타남
 - 적합모형에 변화가 발생한 품목이 모두, 인공신경망모형이 우수한 것으로 판명된 ‘상승 둔화형’과 ‘정체 또는 불규칙 변동형’에서 나타났다는 점은 회귀모형의 변환을 통해 기존 회귀모형의 단점이 어느 정도 보완되었다는 점을 시사함

- 물동량이 상승할 때는 회귀모형이, 그렇지 않을 때는 인공신경망모형이 적합한 것으로 판단됨
 - 지속 상승이든 불규칙 상승이든 물동량의 상승이 계속될 경우에는 회귀모형이 전반적으로 적합도 면에서 우수한 것으로 나타남
 - 상승이 둔화되거나 정체, 혹은 추세를 판단할 수 없을 정도의 물동량 변화를 보이는 품목은 인공신경망모형이 보다 타당한 결과를 낼 수 있을 것으로 판단됨

제6장 몬테카를로 시뮬레이션 분석

- 몬테카를로 시뮬레이션을 이용하여 가상의 자료를 통하여 자료의 예측 값에 대한 변동성을 측정함으로써 신뢰성과 강건성을 평가
 - 시뮬레이션은 회귀분석모형과 인공신경망모형을 대상으로 실시되었으며, 모형 예측의 신뢰성 및 강건성은 20개 품목별 및 4개로 분류된 유형별로 비교
- 시뮬레이션은 시간이 경과함에 따라 오차항 분산이 증가하게끔 설계하였으며, 가상으로 생성된 변수에 의한 추정치를 100번 반복하여 평균값을 구한 후, 이 과정을 1,000번 반복하여 신뢰구간으로 이용
- 신뢰성의 경우, '상승 지속형'은 회귀분석이, 나머지 유형은 인공신경망모형이 높게 나옴
 - 예측 결과의 진폭이 작을수록 예측의 신뢰성은 높으므로 신뢰성은 각 시점별 추정치의 표준편차의 합을 이용하여 측정
 - 1983년에서 2004년의 데이터를 이용하여 신뢰성을 비교
 - '상승 지속형'의 경우 회귀분석이, '상승 둔화형'과 '불규칙 상승형', '정체 또는 불규칙 변동형'은 인공신경망모형의 신뢰성이 더 높게 나타남

- 강건성 또한 신뢰성과 마찬가지로, ‘상승 지속형’은 회귀분석이, 나머지 유형은 인공신경망이 보다 강건한 것으로 나타남
 - 강건성은 모형이 얼마나 일정한 형태를 유지할 수 있는가의 문제이므로, 이상치가 발생하였을 때 신뢰구간이 확장되는 정도를 강건성의 척도로 이용
 - 신뢰성의 경우와 마찬가지로 1983년에서 2004년까지의 데이터를 이용하였으며, 본 연구에서 이상치 발생시점은 1995년으로 설정
 - ‘상승 지속형’에서는 회귀분석, ‘상승 둔화형’, ‘불규칙 상승형’, ‘정체 또는 불규칙 변동형’에서는 신경망 분석이 더 강건한 것으로 나타남

제7장 결론 및 정책대안

- 상승이 지속되는 품목의 경우 전반적으로 회귀분석의 적합도, 신뢰성, 강건성이 우수하게 나타남
 - 지속 상승이든 불규칙 상승이든 물동량의 상승이 지속되고 있는 것으로 평가된 품목의 경우 선형관계를 바탕으로 한 회귀모형이 전반적으로 적합도 면에서 우수한 것으로 나타남
 - 반면 상승이 둔화되거나 정체 혹은 추세를 판단할 수 없을 정도의 물동량 변화를 보이고 있는 품목은 인공신경망모형이 보다 타당한 결과를 내고 있음
 - 각 유형별로 적합도에서 우수한 것으로 판명된 방법의 신뢰성과 강건성이 대부분 더 우수한 것으로 나타남
- 이러한 결과는 기존의 연구이론과도 부합하는 것임
 - 홍콩 항의 항만물동량 예측에 있어 신뢰성과 정확성 측면에서 인공신경망모형이 회귀모형보다 전반적으로 나은 것으로 나타남 Lam, W. H. K. and Pan L. P. and Eddie C. M(2004)의 연구결과와 부합
 - 짧은 기간 동안 등락폭이 심하게 발생하는 품목의 물동량 예측시 인공신경망모형의 우수성을 입증한 신성환(1995)의 이론과도 부합

- 최근처럼 대내외적인 물류여건이 급변하는 시기에는 장기적인 예측뿐 아니라 단기적인 예측도 정기적으로 시행할 필요가 있음
 - 항만물동량에 영향을 미칠 수 있는 대내외적 요인의 변동성을 지속적으로 반영하고, 나아가 장기적인 항만개발계획까지 융통성 있게 수정할 수 있기 위해서는 물동량에 대한 단기적인 예측이 정기적으로 이루어져야 함
 - 이용 가능한 데이터나 항만물동의 특성을 고려할 때 단기에측은 월별 혹은 분기별 단위의 자료를 바탕으로, 분기별 혹은 반기별 예측을 시행하는 것이 바람직할 것으로 판단됨
- 단기 예측은 그 자체의 정확도에 의미를 두는 것이 바람직하므로, 인공신경망모형이 요긴하게 이용될 수 있음
 - 물동량의 변화는 다양한 요인이 복합적으로 작용하여 나타나는 현상으로 특정 요인의 변화가 단기적으로 미치는 정도를 구체적으로 파악하기가 어려우므로, 단기 예측은 그 자체의 정확도에 의미를 두는 것이 바람직함
 - 예측의 정확도가 중시되는 단기적인 예측에는, 모형을 해석하기 어렵다는 단점이 있지만 예측의 정확도가 높은 인공신경망모형이 요긴하게 사용될 수 있음
 - 자료가 축적될수록 인공신경망모형의 효용성은 더욱 높아질 것임
- 장기적인 예측에 인공신경망모형을 도입하기 위해서는 불해석력이라는 단점을 해결해야 함
 - 물동량의 변화에 대처한 정책적인 판단에 도움이 되기 위해서는 물동량에 영향을 미치는 변수와의 인과관계 파악을 위한 정형화된 예측함수가 요구될 수 있음
 - 장기적인 예측은 단기와는 달리 국가경제발전 및 정책방향과 궤를 같이 할 수밖에 없으며 따라서 이를 반영할 수 있는 구체적 함수모형의 설정이 더욱 중요시됨

- 장기적인 예측을 위하여 본 연구에서는 인공신경망모형과 회귀분석의 통합 혹은 겸용 사용을 제안함
 - 품목별 특성에 따라 인공신경망모형의 예측이 유리한 것으로 판단될 경우 우선 인공신경망모형을 이용한 예측치를 산출한 다음, 회귀모형을 적절히 변형시켜 동 예측 결과에 가장 근접한 모형을 최종 예측모형으로 설정하는 방법을 본 연구에서는 제안
 - 인공신경망모형을 통한 예측 결과를 기준으로 삼고 이와 유사한 결과를 얻도록 기본적인 회귀모형의 작위적 변형을 가하는 것임
 - 작위적 모형의 설정은 설명변수와 독립변수와의 관계에 적절한 것으로 평가되는 함수적 변환만을 고려
 - 최종 예측 결과는 물론 선택된 회귀모형에서 도출된 것으로 함
- 향만물동량에 대한 신경망모형의 효용 가능성은 통계적 방법을 통해 검증하였으나, 실제 적용은 모형을 구조적으로 설명해 줄 수 있는 함수가 없는 문제점을 어떻게 해결하느냐에 달려 있음

제1장

서론

1. 연구의 필요성 및 목적

1) 연구의 필요성

2004년 기준 우리나라의 항만물동량은 총 967백만 톤(2005년은 10억 1,100만 톤)에 이르고 있으며, 컨테이너물동량은 1,429만TEU(1,494만TEU)로 규모면에서 중국, 미국, 싱가포르, 일본에 이어 세계 5위¹⁾를 차지하고 있다. 또한 이러한 물동량을 처리하기 위해 그에 걸맞는 항만시설을 갖추고 있으며, 매년 늘어나는 물동량에 대비한 적절한 규모의 항만개발을 위해 수천억 원의 정부재정 및 민간자본이 투입되고 있다. 항만개발에 소요되는 기간이 비교적 긴 편인데다 신항의 경우 10~20년이란 장기간 후를 내다보고 항만개발계획을 수립한다는 점에서 계획수립의 기본이 되는 물동량 수요 전망의 중요성은 아무리 강조해도 지나침이 없을 것이다.

물동량의 수요를 정확히 예측하기 위해서는 우선 물동량의 추세적 변화와 이에 영향을 미치는 대내외적 요인을 정확히 파악하고 이를 적절히 반영해 줄 수 있는 예측모델을 설정하는 것이 가장 중요하다.

'90년대 이후 연평균 두 자릿수의 급증세를 보여 왔던 우리나라 컨테이너물

1) T&F Informa UK Ltd, "World Container Port Traffic League", *Containerisation Yearbook 2006*, 2006.

동량의 경우 2000년대 초반 이후 증가세가 급격히 둔화되는 조짐을 보이고 있다. 컨테이너물동량의 증가세는 2004년 10.2%, 2005년 4.4%의 증가에 그쳤고, 특히 매년 20%대 이상의 물동량 증가세를 보여 왔던 환적물동량의 증가율은 각각 12.18%와 7.14%로 크게 감소하는 추세를 보이고 있다.

이러한 증가세 둔화의 원인은 대내적으로 전통 제조업 성장세의 둔화와 첨단 신산업의 부상, 서비스 산업의 비중 확대 등과 대외적으로는 양산 항을 비롯한 중국의 대규모 항만개발, 중국의 다국적 물류기지화 등 동북아 물류체제 개편 현상 등에서 찾을 수 있다. 이 밖에 국내 제조공장의 해외 이전 확대와 FTA(자유무역협정)의 확산, 경제자유구역지정 등 우리에게 긍정적이든 부정적이든 물동량의 흐름에 영향을 미칠 수 있는 주변 여건이 크게 변화하고 있다.

이 같은 경제·사회적 여건의 급속한 변화를 고려할 때 주어진 과거 데이터의 추세적 변화요인을 지나치게 증시하는 기존 항만물동량 예측방법²⁾의 구조적 한계를 벗어날 필요가 있다. 즉, 최근 나타나고 있는 기술혁신과 산업구조조정에 따른 물동량 특성의 변화를 제대로 반영할 수 있는 보다 다양한 예측방법론을 개발할 필요성이 제기되고 있는 것이다.

2) 연구의 목적

현재 전국 단위의 항만물동량 예측은 해양수산부의 ‘전국 무역항 기본계획’ 수립을 위해 5년 단위로 이루어지고 있으며, 예측의 초점은 10~20년 단위의 중장기에 맞추어져 있다. 그러나 지금과 같이 내·외부의 물류환경이 급변하는 상황에서 중장기적인 예측이 정확할 것을 기대하기는 어렵다. 항만개발계획이 포괄적이고도 장기적인 관점 하에 수립되어야 함은 당연하나, 수요 예측에 대한 지속적인 보완 및 수정이 따르지 못할 경우 자칫 엄청난 재정적 낭비를 초

2) 기존에 주로 사용되는 항만물동량 예측방법으로는 회귀모형을 다양하게 변형시킨 변형회귀모형(transformed regression model) 혹은 시계열모형(time series model)이 있음. 이들 모형만으로는 품목별 물동량 예측의 정확도(precision) 및 신뢰도(reliability)를 크게 향상시키기에 한계가 있음.

래할 수 있다. 이에 정부에서는 항만물동량의 변동성 감지 및 예측 기능을 전담하는 가칭 ‘수요예측센터’를 본 원에 설립하여, 정부의 장기 항만개발 기본계획을 탄력적으로 조정·운영하려 하고 있다. 따라서 연도별 실적자료를 바탕으로 한 중장기적 예측뿐만 아니라 월별 혹은 분기별 자료를 이용한 1~2년 단위의 단기적인 예측이 실시간으로 행해질 필요가 있다. 월별 혹은 분기별 자료를 이용한 단기적인 예측은 데이터의 용량이 클 뿐더러 그만큼 정확성이 요구된다. 이런 점에서 데이터의 일정한 추세를 전제로 중장기적인 시점을 예측해 왔던 회귀분석(Regression)과 구분될 수 있는 보다 정교한 예측모형의 사용 필요성이 제기되고 있다.

본 연구는 최근 다양한 분야에서 주목받고 있는 데이터 마이닝(Data Mining) 기법에 의한 수리 예측방법의 활용에 초점을 맞추고 있다. 즉, 이러한 수리예측방법을 항만물동량에 적용시켜 이를 예측해 보고, 예측의 신뢰도 및 정확도를 비롯한 다양한 통계적 특성들을 기존 모형과 비교해 봄으로써 동 기법의 유용성을 평가하고자 하였다.

특히 다양한 데이터 마이닝 기법들 가운데 일반적으로 많이 이용되는 신경망분석모형(Neural Network Model ; NNM)에 주목하였으며, 몬테카를로 시뮬레이션을 이용하여 모형 적합성에 대한 객관적 신뢰도 분석을 수행하였다.

본 연구의 목적은 주요 품목별 물동량의 추세적 특성과 연관해 그에 적합한 예측모형의 적용 가능성을 검토함으로써 궁극적으로는 물동량 예측 결과의 신뢰도를 향상시키는 데 있다. 이를 위해 기존에 사용되어 왔던 회귀모형과 시계열모형 외에 적용 가능한 새로운 예측기법으로서 데이터 마이닝 기법의 대표적인 예측모형인 신경망분석모형의 활용방안을 강구하고자 한다.

2. 연구 범위와 방법

인공신경망모형의 항만물동량 적용 가능성 여부에 대한 판단은 기존에 사용되어 왔던 예측모형과의 비교·분석 하에 이루어지며 그 구체적 활용방안은 비교·분석 과정에서 드러난 특징과 모형 자체의 장단점을 고려하여 설정되어야 한다. 따라서 본 연구의 범위는 기존 예측모형의 개선 대안으로서 제시된 인공신경망모형의 유용성을 판단하고 최적 활용방안을 모색하기 위한 일련의 과정으로 정의할 수 있다.

기존에 사용되어 온 예측모형은 일반적인 다중회귀분석에 대부분 기초하고 있으며, 모형의 설득력을 높이기 위해 설명변수 혹은 독립변수를 여러 함수형태로 변환시킨 변형회귀모형도 많이 사용되고 있다. 항만물동량의 예측이 통상 품목별로 이루어지기 때문에 품목별 물동량의 추세적 특징에 따라 회귀모형을 다양한 형태로 변환시켜 사용함을 의미한다. 물론 인공신경망모형 자체도 예측 대상의 추세적 특징에 따라 예측의 정확도에 차이가 날 확률이 높기 때문에 기존 예측모형과의 비교·분석은 당연히 품목별로 이루어져야 한다.

본 연구에서는 예측모형 간 비교 대상으로 수출과 수입 분야에서 각각 금액 기준 상위 10개 품목을 선정하되 항만에서의 안벽시설이 요구되지 않는 유류 품목은 제외하였다. 신경망모형과의 비교·분석은 기본적인 회귀모형과 변형된 회귀모형 두 가지 모두에 대해 실시하였다. 예측을 위한 설명변수의 설정은 이미 상기 20개 품목에 각각에 대한 예측이 여러 번 반복 시행되어온 점을 고려해 기존에 사용되었던 변수들을 대상으로 요인분석 과정을 거쳐 최종 확정하였다. 비교 대상인 변형회귀모형의 경우 품목별로 설정된 변수에 대한 다양한 형태의 함수변환이 가능할 수 있으나, 본 연구에서는 당 원에서 기 시행하였던 예측모형을 비교대상으로 설정하였다.

품목별 예측 결과에 대한 비교우위성 판단은 예측의 적합도(fitness), 안정성(reliability), 강건성(robustness)을 바탕으로 하였다. 단, 모형의 유용성 및 그

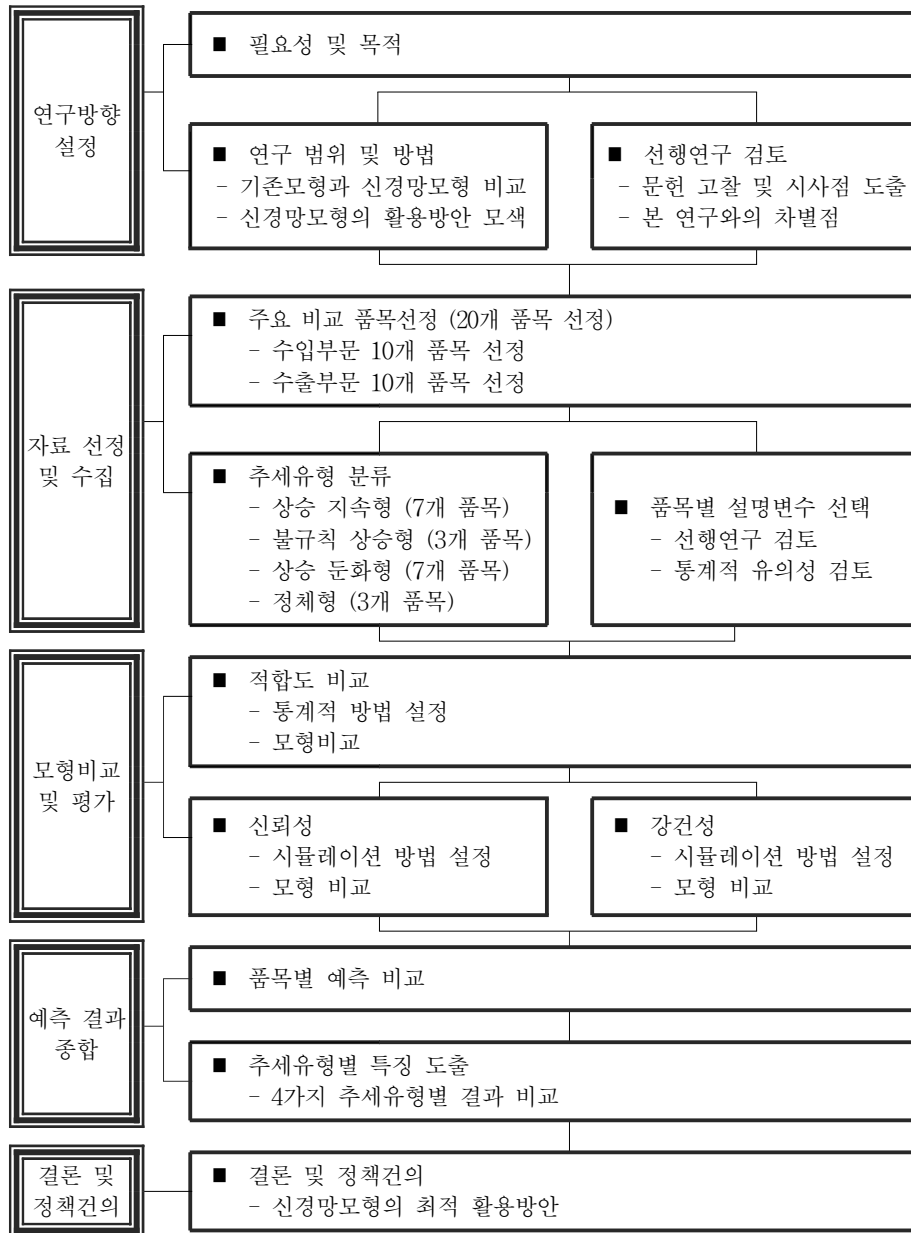
특징을 판단하기 위해 품목별 물동량의 변화유형을 ‘성장 지속형’, ‘불규칙 성장형’, ‘성장 둔화형’, ‘정체 혹은 불규칙형’ 등 크게 4가지 범주로 분류하였다. 이것은 모형의 적합성 여부를 개별 품목별로 판단하되 비교우위에 대한 종합적 분석 및 판단을 4가지 변화유형과 연계시킴으로써 물동량의 변화추세에 부합될 수 있는 신경망모형의 적절한 예측 활용방안을 모색하기 위해서이다. 회귀분석모형과 신경망모형의 신뢰도 및 강건성 분석은 몬테카를로 시뮬레이션(Monte Carlo Simulation) 방법을 이용하였다.

본 연구는 모두 7장으로 구성되어 있다. 제2장에서는 이번 연구에 대한 이해를 돕기 위해 데이터 마이닝 기법에 대한 개념과 특징, 그것의 폭넓은 활용 분야에 대해 간단히 언급하였다. 또한 대표적인 예측기법으로 인공신경망모형의 특성과 기타 예측방법과의 차이점을 설명하였다. 제3장에서는 국내외 선행 연구를 선형모형과의 비교연구 및 시계열모형과의 비교연구, 그리고 기타 관련 연구로 구분하여 고찰하였고 본 연구와의 차별성을 설명하였다.

제4장에서는 모형간 비교·분석을 수행하기 위해 주요 품목의 선정과정을 기술하였고, 각 품목별 항만물동량의 추세 변화의 특징을 4가지 유형으로 분류하였다. 이와 함께 개별 품목의 모형 간 적합도의 비교·분석을 통해 4가지로 구분된 물동량 추세변화의 유형과 관련한 공통적인 특징을 도출하고 있다. 제6장에서는 몬테카를로 시뮬레이션을 이용하여 회귀모형 및 신경망모형 간 신뢰도 및 강건성을 비교하였고, 4가지 물동량 변화 유형별로 제5장에서 비교된 적합도 결과와의 차이점 여부를 분석하였다.

제7장에서는 연구결과를 종합하고 신경망분석모형의 최적 활용방안을 모색하였다. 본 연구의 추진절차는 <그림 1-1>과 같다.

〈그림 1-1〉 연구 추진 흐름도



제2장

데이터 마이닝 기법에 대한 정의

1. 데이터 마이닝의 개념

데이터 마이닝은 일반적으로 데이터베이스에 있는 방대한 양의 정보에 숨어 있는 지식을 자동적으로 추출하는 과정이다. 데이터 마이닝은 방대한 데이터베이스로부터 숨어있는 예측정보를 추출하는 것이라고 설명할 수 있다.³⁾ 자료로부터 암시적이고 사전에 알려지지 않은, 유용한 정보를 추출하는 것이다. 정보나 의사결정을 식별하기 위해서 다양한 기술들을 사용하는 것으로 자료들로부터 지식을 추출하여 의사결정 지원, 예측, 추정과 같은 영역에 사용될 수 있도록 하는 것이 데이터 마이닝의 기본 개념이라 할 수 있다. 이러한 정보는 예를 들어, 추세 또는 패턴 등이며 중요 비즈니스 프로세스 개선에 사용될 수 있다.

데이터 마이닝은 인공지능 분야의 기계학습⁴⁾ 이론에 그 뿌리를 두고 있는데 기계학습은 규칙을 찾아내기 위한 자동화된 유도과정이라 할 수 있다. 기계학습에서는 트레이닝 셋이라 불리는 작은 양의 실험실용 데이터를 사용하여 알고리즘을 만들어 내지만 데이터 마이닝은 현실 세계의 대규모 데이터베이스를 트레이닝 셋으로 간주해서 이로부터 유용한 지식을 캐내는 작업을 수행한다. 데이터베이스로부터 과거에는 알지 못했지만 데이터 속에서 유도된 새로운 데이터 모델을 발견하여 미래에 실행 가능한 정보를 추출해 내고 의사 결정에

3) R. Roiger and M. Heatz, *Data Mining: A Tutorial Based Primer*, Addison Wesley, 2003.

4) 기계학습(Machine Learning)은 주로 전산과학에서 사용되는 용어로 통계학에서는 통계학습(Statistical Learning)이라 불리기도 한다.

이용하는 과정을 말하는 것이다. 즉 데이터에 숨겨진 패턴과 관계를 찾아내어 광맥을 찾아내듯이 정보를 발견해 내는 것이다. 여기에서 정보 발견이란 데이터에 고급 통계 분석과 모델링 기법을 적용하여 유용한 패턴과 관계를 찾아내는 과정이다.

1) 데이터 마이닝의 기능

데이터 마이닝의 목적은 사용자에게 유용한 새로운 지식을 생성하는 것으로 현실세계에 적합한 모형의 수립이 전제되어야 한다. 수립된 모형은 데이터에 내재된 패턴이나 관계를 설명하는 목적으로 다음의 두 가지 종류로 나눌 수 있다.⁵⁾

(1) 서술모형

결과가 알려져 있지 않은 데이터를 가지고 예측하는 방식이다. 즉, 데이터 내의 숨겨진 관련성을 찾는 것을 목적으로 한다. 이 방식에는 군집분석, 연관성 분석, 서술분석 등이 속한다.

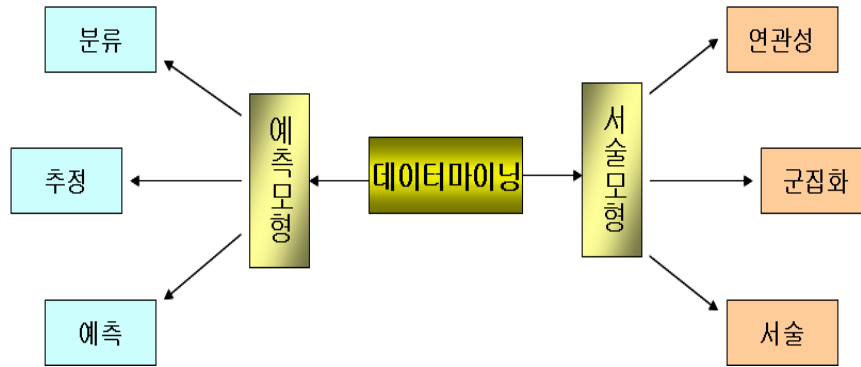
(2) 예측모형

이미 알려진 발견된 결과를 사용하여 모형을 구축하고 이를 응용하여 미래를 예측하는 방식이다. 따라서 과거 데이터를 분석하여 모형을 만들어야 한다. 이 방식에는 분류, 추정, 예측 등이 속한다.

데이터 마이닝 모형을 종류별로 정리한 그림은 아래와 같다.

5) 예측모형은 목표변수가 존재하는가, 그렇지 않은가에 따라 지도예측(Supervised Prediction)과 자율예측(Unsupervised Prediction)로 구분되는 경우도 있다.

〈그림 2-1〉 데이터 마이닝 모형의 종류



2) 데이터 마이닝 응용분야

데이터 마이닝 기법이 현재 가장 많이 사용되고 있는 분야는 데이터베이스 마케팅으로 그 분야의 핵심 기술이 되어가고 있다. 데이터 마이닝을 이용한 다른 분야에의 응용은 아직은 초기 단계에 머물러 있다. 그러나 곧 일반화될 것으로 생각된다. 그 이유는 데이터 마이닝을 하기 위한 최적의 시스템이 되는 데이터 웨어하우스가 국내에 이미 많이 구축되어 있기 때문이다. 또한 기술적인 면에서 데이터 마이닝을 잘 구현할 수 있는 소프트웨어가 출시되고 있기 때문에 사용상의 제약이 경감되고 있기 때문이다. 지금까지 알려져 있는 데이터 마이닝 기법의 응용분야는 다음과 같다.

(1) 마케팅⁶⁾

데이터 마이닝이 가장 많이 사용되는 분야가 마케팅이다. 기업이 가지고 있는 고객 정보의 데이터베이스를 중심으로 데이터 마이닝 기법을 사용하여 데

6) 데이터 마이닝이 많이 활용되는 마케팅 분야로 고객관계기법(Customer Relationship Management)을 예로 들 수 있다.

이터베이스 마케팅에서의 주제에 응용되고 있다. 첫째로, 고객의 인구 통계 자료나 구매패턴의 정보를 기반으로 고객을 세분화하고 그 특성을 요약하여, 그 결과를 바탕으로 Target 마케팅에 활용함으로써 적은 비용으로 최대의 효과를 얻기 위해 활용할 수 있다. 둘째, 고객들의 성향을 파악하여, 경쟁 업체로의 전환 가능성이 있는 고객 혹은 더 이상 자사의 제품이나 서비스를 이용하지 않는 고객들을 분류하여 고객 유지율을 향상시키고, 이탈한 고객들을 다시 자사의 고객으로 되돌릴 수 있는 방안을 모색함으로써 고객과의 지속적인 관계를 유지해나가는 데 이용한다. 셋째로, 마케팅 활동의 결과에 대한 효과분석을 함으로써 데이터베이스 마케팅에 활용할 수 있다.

(2) 소매업⁷⁾

가장 큰 활용 범위로는 시장바구니 분석이 있다. 즉, 생산품들의 각 구매 시점을 이용하여 생산품 간의 관련성을 찾아내고, 이익을 극대화하기 위한 전략을 세우는 데 활용이 된다. 활용의 예로는 일시적인 유행이나 구매 행동을 읽을 수 있는 순차적인 패턴을 탐지하여 판매 전략을 세우는 것 등이다.

(3) 금융업

잠재된 신용카드의 도용 패턴을 탐색해 내거나 이자율이나 유통시장에서의 환율변동 예측에 데이터 마이닝 기법이 이용되고 있다. 또한, 주식과 채권의 포트폴리오 관리에도 응용되고 있으며, 신용 위험 관리와 기업대출에서의 도산 예측시에도 데이터 마이닝 기법으로 훌륭한 결과를 얻고 있다.

(4) 물류/교통

미래 물류량을 예측하여 수요에 맞는 인프라를 구축하는 계획을 시기적절하게 세움으로써 예산과 인력의 효율적 사용을 확보할 수 있다. 또한 높은 신뢰

7) Paolo Giudici, *Applied Data Mining: Statistical Methods for Business and Industry*, John Wiley&Sons Ltd., 2003.

도와 적시 배달이 가능한 분배 포인트를 파악하거나 소매점에서의 분배 스케줄에서 공통의 패턴을 파악, 배송물을 통합하여 소매점들 사이에서 운송 스케줄을 결정한다. 또는 하역 패턴을 분석할 수도 있다.

2. 데이터 마이닝 분석 절차⁸⁾

1) 샘플링/선택

샘플링이란 방대한 양의 데이터(모집단)에서 모집단을 닮은 적은 양의 데이터(샘플, 표본)를 추출하는 것이다. 이는 데이터 마이닝 작업을 하는 데 있어서 시간과 비용을 절감하여 효율적인 작업 진행을 돕고, 이를 통하여 데이터 마이닝 작업의 무게를 조금 가볍게 할 수 있는 절차이다. 그러므로 샘플링이 적절한가에 대한 질문을 던져보고, 필요하다면 적절한 방법을 이용하여 샘플 데이터를 생성한다. 샘플링 방법에서 가장 유의할 점은 모집단을 닮지 않고 한 쪽으로 치우쳐진 샘플이 추출되는 경우로서, 이는 샘플링 후 기본적인 몇 가지의 자료탐색을 거쳐 반드시 확인하여야 한다.

2) 탐색

데이터의 탐색과정에서는 이미 알고 있는 사실들을 확인하여 수치화하는 작업을 시작으로 하여 보유하고 있는 수많은 변수들의 관계를 살펴보는 단계이다. 또한 이 단계에서는 각 변수에 대한 분포, 데이터 마이닝 과제에서 설정한 주제와의 관계 및 변수 간의 선형, 비선형 관계 등을 관찰하게 된다. 이 과정에서는 분산 분석(ANOVA)이나 교차표 등의 통계분석 방법이 사용되기도 하지만 각종 그래프를 통한 시각적 관찰이 유용하게 쓰인다. 특히 시각적 관찰을

8) 이 절차를 SEMMA(Sampling, Exploration, Modification, Modeling, Assessment)라고 칭하기도 한다.

위한 방법들이 다양화되고 점단의 알고리즘에 의해 구현되어 최근 자료 분석에 많은 비중을 차지하고 있다. 이 과정에서 모르고 있던 정보 또는 기존 현업에 있는 사람들이 확신하고 있던 사실들이 틀렸다는 것도 확인할 수 있다.

이처럼 탐색 단계에서는 데이터 모양을 면밀히 둘러보아 정보화할 수 있는 기반을 잡아가는 과정이다. 이 단계에서는 실제 한 모집단으로 알고 있던 것이 두 개 이상으로 나뉠 수도 있고, 수십 개 이상의 많은 변수를 실제 중요한 정보를 주는 소수의 변수로 축소시킬 수도 있으며, 기존에 있던 변수들 외에 새로운 변수를 생성해야 할 필요성을 느끼게 될 수도 있다.

3) 변형과 조정

데이터의 변형은 기존의 변수를 이용하여 새로운 변수를 얻기 위해 이루어진다. 기존의 변수를 이용하여 새로운 변수가 생성될 수 있는데 이러한 작업을 고려하는 단계가 바로 변형 및 조정 단계이다. 이 단계에서 생성 또는 수정된 변수는 차후 모형화 단계에서 아주 중요한 정보로 활용될 수 있을 것이다.

4) 모형화

데이터 마이닝 과정에서 가장 중요한 단계로서, 앞서 선행되었던 단계에서 선정된 주요한 변수를 사용하여 다양한 모형⁹⁾(인공신경망, 의사결정나무, 일반화선형모형 등 통계모형)을 접합해 보는 단계이다. 잘 알려져 있는 모형이 있기도 하나, 그렇지 않은 경우에는 다양한 모형에 적합한 후 예측력이 가장 뛰어난 모형을 선택하는 것이 일반적이다. 데이터 마이닝에서는 대부분 비모수적인 모형을 사용한다. 비모수적인 모형의 단점을 극복하기 위해 분석에 사용할 전체 자료 중 70%만 모형 적합에 사용하고 나머지는 얻어진 모형의 정확성

9) I. Witten & E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, 2000.

을 측정하기 위한 확인 자료(validation data)로 사용한다. 여러 모형들에 대한 평가는 동일한 확인 자료에 의한 오류에 의해 이루어지고 그 중 가장 정확하고 안정적이며 목적에 맞는 모형이 선택되어 현실에 응용된다.

5) 평가

적합한 두 개 이상의 모형 효과를 비교하여 가장 좋은 모형을 선택할 수 있다. 이렇게 선택된 모형은 실제 모집단에 반영하여 그 효과를 재평가할 수 있으며 이로 인하여 새로운 문제를 제거하여 이전 작업을 다시 반복할 수도 있을 것이다.

3. 데이터 마이닝 기법 소개

1) 기술모형

(1) 군집분석

데이터 마이닝에 있어서 분석과정 중, 통계적 기법에 속하는 방법이다. 군집 분석은 다양한 특성을 지닌 관찰 대상을 유사성을 바탕으로 동질적인 집단으로 분류하는 데 쓰이는 기법이다. 분류규칙이 불분명하거나 또는 집단의 수를 미리 정하지 않는 경우에는 군집분석이 매우 유용하다. 전체 집단을 몇 개의 의미 있는 집단인 군집으로 분류하는 데 이용되는 방법인 군집분석은 나누는 대상인 표본크기에 따라 두 가지¹⁰⁾로 나뉜다. 간단히, 표본수가 적은 경우(예를 들어 20~30개 이하)와 표본의 수가 많은 경우(예를 들어 100개 이상)에 따라 다르다. 대개 표본수가 적은 경우에는 계층적 군집방법을, 표본수가 많은

10) R. Johnson & D. Wichern, *Applied Multivariate Statistical Analysis*, 5th Edition, Prentice Hall, 2002.

경우를 K-Means 군집방법을 사용하고 있지만 실제로 적용되는 경우를 보면 이와는 상당히 다르다고 볼 수 있다.

대용량의 데이터가 너무 복잡할 때에는 이를 구성하고 있는 몇 개의 군집을 우선 살펴봄으로써 전체에 대한 윤곽을 잡을 수 있을 것이다. 군집분석은 이러한 상황에 유용하게 쓰일 수 있다.

(2) 연관성 분석

연관성 분석은 어떤 특정 문제에 대한 예측이나 고객들을 특정 목적에 따라 분류하는 문제가 아니라, 상품 혹은 서비스의 거래기록 데이터로부터 상품 간의 연관성 정도를 측정하여 연관성이 많은 상품들을 그룹화하는 군집화의 일종으로서, 동시에 구매될 가능성이 큰 상품들을 찾아냄으로써 시장바구니 분석¹¹⁾에서 다루는 문제들에 적용될 수 있다. 연관성 측정에서 얻어지는 결과물인 연관 규칙은 ‘If A, then B(A \rightarrow B)’ 와 같은 형식으로 표현된다. 연관성 측정에서의 연관 규칙은 ‘상품 A가 구매된 경우 상품 B도 구매된다’ 라고 해석할 수 있다. 연관규칙은 구체적인 상품들이 언급되므로 이해가 쉽고 명쾌하며 실제 업무로의 적용이 용이하다. 즉, 유용한 규칙을 찾게 된다면 상당히 실질적이고 효과적인 정보를 얻을 수 있게 되는 것이다.

2) 예측모형

(1) 의사결정나무¹²⁾

의사결정나무는 분류 및 예측에 있어서 자주 쓰이는 기법으로, 통계학적인 용어를 쓰지 않고도 사건에 영향을 미치는 변수들과 변수들의 상호작용을 쉽게 설명할 수 있다는 것이 장점이다. 또한, 의사결정나무는 인공지능망과 달리

11) M. Berry & G. Linoff, *Data Mining Techniques*, John Wiley & Sons, Inc., 1997.

12) L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*, Wadsworth, 1984.

이해하기가 쉬워 널리 사용되고 있다. 그 외 이 방법의 장점은 다음과 같다.

- ① 분류나 예측의 근거를 알려주기 때문에 이해하기가 쉽다.
- ② 연속형 명목형 데이터 모두 사용 가능하다.
- ③ 어떠한 속성들이 각각의 예측값에 결정적인 영향을 주는가를 쉽게 파악할 수 있다.
- ④ 모형구축에 소요되는 시간이 짧다.

하지만 이는 여러 변수를 조합한 경우에는 규칙을 발견할 수 없고 가지 사이의 분할이 매끄럽지 못하여 분할을 선택할 때 분할 결과가 장래의 분할에 어떤 결과를 초래할지 알 수 없다는 단점이 있다. 또한 예측의 경우에 예측력이 인공신경망에 비해 떨어질 수 있다는 단점도 있다. 그 외 이 방법의 단점은 다음과 같다.

- ① 연속형 데이터를 처리하는 능력이 인공신경망이나 통계기법에 비해 떨어지며, 결과적으로 예측력도 감소한다.
- ② 모형을 구축하는 데 사용되는 표본의 크기에 지나치게 민감하다.

(2) 인공신경망모형¹³⁾

인공신경망모형은 데이터 마이닝에 대한 관심이 모아지면서 가장 일반적으로 언급되고, 또한 다양한 응용 분야를 가지고 있는 기법이다. 인간 두뇌의 신경세포를 모방한 개념으로 마디(node)와 고리(link)로 구성된 망구조를 모형화하고, 의사결정나무와 마찬가지로 과거에 수집된 데이터로부터 반복적인 학습 과정을 거쳐 데이터에 내재되어 있는 패턴을 찾아내는 모델링 기법이다. 분류, 군집, 연관규칙 발견과 같은 작업에 널리 사용된다. 인공신경망모형은 인간이 경험으로부터 학습해 가는 두뇌의 신경망 활동을 흉내 내어 자신이 가진 데이터로부터 반복적인 학습과정을 거쳐 패턴을 찾아내고 이를 일반화함으로써 특

13) B. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.

히 향후를 예측(Prediction)하고자 하는 문제에 유용하게 쓰인다.

매우 복잡한 구조를 가진 데이터들 사이의 관계나 패턴을 찾아내는 유연한 비선형모형(Flexible nonlinear Model)의 하나로, 신경생리학과 유사성 때문에 일반적으로 다른 (통계적) 예측모형에 비해 흥미롭게 여겨지고 있다. 이러한 인공신경망모형은 고객의 신용평가, 불량거래의 색출, 의료진단 예측, 우량 고객 선정, 타겟 마케팅의 여러 주제 등을 비롯한 여러 분야에 적용될 수 있는데, 주로 지도학습(supervised learning)에 적용되어 목적변수에 대한 예측이나 분류를 목적으로 감춰진 패턴을 찾고 이를 일반화하는 데 이용된다. 혹은 비지도학습(unsupervised learning)에서 코흐넨 맵(Kohonen maps)¹⁴⁾을 이용하여 데이터의 군집화 작업을 수행하는 데 쓰이기도 한다.

인공신경망모형은 상당히 다양한 산업분야의 여러 문제에 적용될 수 있고, 입력변수와 목적변수의 관계를 그리기가 어려운 복잡한 데이터에 대해서도 좋은 결과를 주는 것으로 알려져 있다. 또한 입력변수와 목적변수의 속성이 연속형이나 이산형인 경우를 모두 다룰 수 있다. 그러나 인공신경망모형은 매우 유연하기는 하지만 설명력이 부족하여 종종 블랙박스로 불리고 있다. 의사결정 나무나 연관성 분석은 명쾌하고 쉽게 이해할 수 있는 결과물(일종의 rule)을 제공하는 데 반해, 인공신경망모형은 인간이 어떠한 현상을 인지하게 되는 것처럼 쉽게 설명되지 않는 내부적인 작업을 수행하고 이를 통해 얻어진 결과물을 제공할 뿐 어떠한 변수가 중요한지, 어떻게 상호작용이 이루어져 그러한 결과물을 주게 되는지에 대한 설명은 하지 않는다. 따라서 설명력보다는 정확한 예측이 중요한 경우에 이용될 수 있다. 인공신경망의 장점은 다음과 같다.

- ① 입력과 출력마디에 이산형, 연속형 변수들을 모두 사용할 수 있다.
- ② 입력변수들의 비선형 조합을 통해 제공하기 때문에 예측력이 우수하다.
- ③ 제품 선택의 폭이 넓고 취득하기 쉽다.

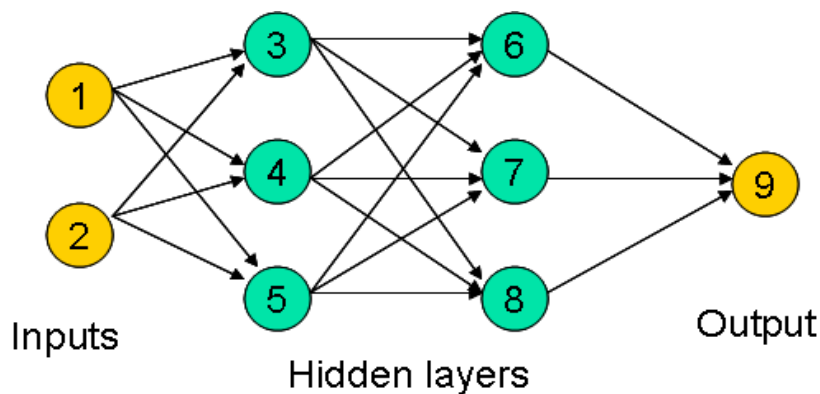
14) T. Kohonen, *Self-organization and Associative Memory*, 3rd edition, Springer-Verlag, Berlin, 1989.

인공신경망의 단점은 다음과 같이 요약해 볼 수 있다.

- ① 결과에 대한 이유를 설명하지 못한다.
- ② 변수 변환의 추가적이 노력이 필요하다.
- ③ 복잡한 학습과정이 필요하므로 모델을 구축하는 데 시간이 많이 걸린다.
- ④ 비전문가들이 사용하기에 어려움이 따른다.

인공신경망모형을 자세히 서술해 보고자 한다. 먼저 인공신경망은 입력층, 은닉층, 출력층으로 구성되어 있다. 입력층은 각 입력변수에 대응되는 노드로 구성되어 있으며 노드의 수는 입력변수의 개수와 같다. 은닉층은 입력층으로부터 전달되는 변수 값들의 선형결합을 비선형함수로 처리하여 출력층 또는 다른 은닉층에 전달하는 역할을 한다. 마지막으로 출력층은 반응변수에 대응되는 노드로 예측값이 생성되는 곳이다. 이를 그림으로 표현하면 다음과 같다.

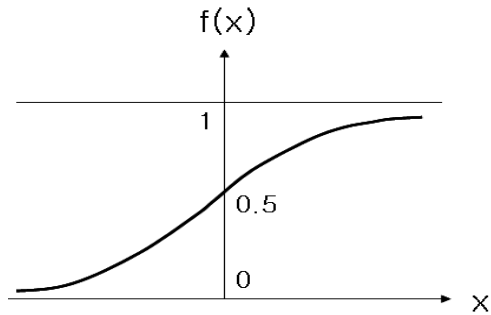
〈그림 2-2〉 인공신경망모형의 구성



위 그림에서 입력노드(1, 2)는 예측변수에 해당하고 제1은닉층 내 은닉노드(3, 4, 5)는 예측변수들의 선형결합이며 제2은닉층 내 은닉노드(6, 7, 8)는 제1은닉 노드들의 활성화함수 값의 선형결합이다. 그리고 출력노드(9)는 제2은닉 노

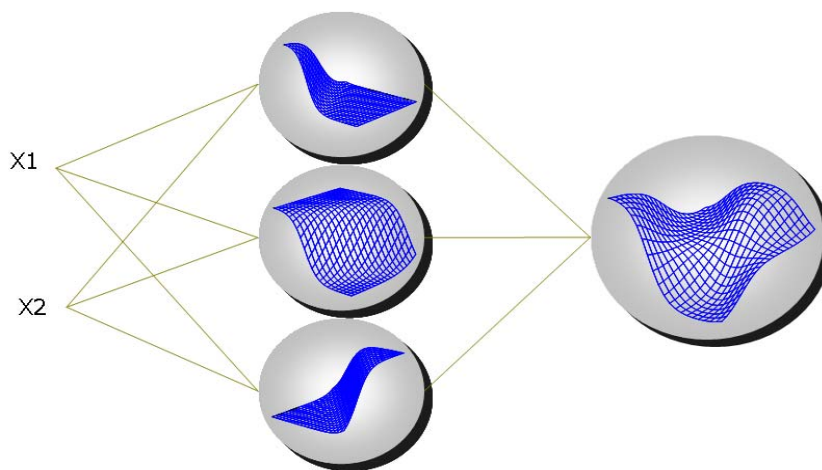
드들의 활성화함수 값으로 통상 예측값을 의미한다. 여기서 활성화함수란 비선형으로 변형을 해주는 함수로 흔히 시그모이드(sigmoid)¹⁵⁾ 함수가 많이 사용된다. 시그모이드 함수란 다음과 같은 S자형 곡선을 의미한다.

〈그림 2-3〉 시그모이드 함수의 형태



이상의 인공신경망의 모형화 과정을 요약해 그림으로 표현하면 다음과 같다.

〈그림 2-4〉 인공신경망모형화의 과정



15) 시그모이드 함수는 로지스틱 함수로 칭해지기도 한다. 그 외에도 쌍곡탄젠트(hyperbolic) 함수를 사용하는 경우도 있다.

즉, 두 개의 입력변수로부터 선형결합을 취한 다음 은닉층에서 활성화함수 값을 구한 후 이를 다시 선형 결합하면 우측과 같은 비선형함수가 구해진다. 이 비선형함수에 의한 예측값은 일반적인 선형함수에 의한 예측값보다 더 정교하다고 알려져 있다.

(3) 기타 통계 예측모형

예측을 위한 통계모형은 오랜 시간에 걸쳐 개발되어 왔다. 그 중 가장 대표적인 예측모형은 회귀모형일 것이다. 그 외에도 로지스틱회귀모형, 판별분석, 일반화 선형모형 등 많은 기법들이 있다.

위에서 소개한 데이터 마이닝 기법들이 실제 응용 사례에서는 주로 어떤 분야에서 사용되는지를 아래 표에 정리하였다.

〈표 2-1〉 데이터 마이닝의 응용분야

응용분야	데이터 마이닝 기능	데이터 마이닝 기법
시장바구니 분석	기술	군집분석, 연관성 분석
타겟 마케팅 부정행위 적발	예측(분류)	의사결정나무, 인공신경망 로지스틱회귀모형, 판별분석
물동량예측 주가예측	예측	회귀분석, 의사결정나무 인공신경망, 일반화선형모형
시장세분화	예측(분류), 기술	인공신경망, 의사결정나무 군집분석, 판별분석

4. 인공신경망의 특성과 차이점

1) 인공신경망의 특성

인공신경망 분석에는 다른 분석방법들이 갖지 못하는 좋은 특성들이 있다.

(1) 병렬처리

기존의 예측방법은 모든 절차를 순차적으로 이루는 직렬형태였다. 이와는 대조적으로 인공신경망은 다수의 노드가 모여서 동시에 서로 다른 처리를 한다. 이 때 노드 각각은 그 처리속도가 느리지만 여러 개에 의한 병렬분산처리로 직렬처리보다 빠른 분석을 수행할 수 있다.

(2) 학습과 훈련을 통한 적응성

인공신경망은 인간의 두뇌처럼 경험을 통하여 학습하는 기능이 있다. 얻어진 자료를 가지고 입력층에서 마지막 출력층에 이를 때까지 다음 층에 가중치를 부과하는 과정을 반복해 가면서 모형을 만들어 분석한다. 실제값과 관측값 사이에 발생하는 차이를 이용하여 학습하고 새로운 모형을 구축하게 된다.

(3) 노드 간의 고도의 상호연결성

인공신경망은 정보를 기억장치의 정해진 장소에 일괄 저장해 두지 않고, 노드라 불리는 많은 처리장치에 저장한다. 이는 노드 간에 고도의 연결성을 의미한다. 이는 하나의 정보가 네트워크 전체나 어느 부분에 걸쳐 분산되어 있을 수도 있으며, 여러 개의 정보가 하나의 네트워크에 동시에 저장될 수도 있다. 이것으로 인하여 일부 정보가 파손된다 하더라도, 작업을 계속 수행할 수 있다.

2) 인공신경망과 기타 예측방법과의 차이점

예측모형으로 인공신경망과 비교될 수 있는 기타 방법들로는 의사결정나무, 회귀분석, 시계열분석, 일반화 선형모형 등이 있다. 기타 방법들은 각각의 장단점이 있으며 인공신경망 분석과 비교하면 다음과 같은 차이가 있다.

(1) 처리가능 데이터의 다양성

인공신경망과 의사결정나무 방법들은 데이터의 값들이 연속형 변수일 때는 예측을 실행하며, 이산형 변수일 때는 분류를 실행한다. 하지만 회귀분석, 시계열분석, 일반화 선형모형 등에는 반드시 연속형 변수만이 사용되어야 한다.

(2) 복잡한 인과관계 자료의 처리능력

인공신경망은 이 부분에서 탁월한 능력을 발휘한다. 인공신경망은 비모수적 비선형함수를 적합하는 방법이므로 복잡한 인과관계가 숨어있는 경우에도 이를 잘 찾아내어 모형화하는 능력이 있다. 반면에 의사결정나무와 회귀분석은 단순한 모형에 적합하므로 복잡한 구조의 데이터에는 적합하지 않다. 일반화 선형모형 역시 모형의 구조가 단순하므로 이 부분에서 취약하다.

(3) 예측력

인공신경망의 최대 장점이다. 기존 의사결정나무와 회귀분석, 그리고 시계열 분석은 예측력에 한계를 가지고 있다. 그 이유는 모형에 선형성을 가정하였기 때문이다. 일반화 선형모형은 비선형성이 있으므로 회귀분석보다는 유연하지만 인공신경망보다는 덜 유연한 모형을 찾게 된다. 따라서 일반화 선형모형의 예측력은 향상되나 인공신경망의 예측력에는 못 미친다.

(4) 해석력

인공신경망의 가장 큰 약점이다. 기존 의사결정나무와 기타 통계방법들은

우수한 해석력을 가지고 있다. 요즘 인공지능망의 해석력을 보완하기 위해 인공지능망 분석 후 의사결정나무를 사용하여 해석을 돕는 방법이 나오고 있다.

(5) 프로그램 접근 및 용이성

인공지능망과 의사결정나무는 많은 통계프로그램¹⁶⁾에 적재되고 있는 중이다. 따라서 기존 회귀분석과 시계열분석 방법들의 접근 용이성에 근접해 있다. 하지만 사용 용이성에 있어서는 시계열분석이나 일반화 선형모형인 경우, 사용자가 통계를 잘 알지 못하는 경우에는 모형의 적합에 어려움이 있을 수 있다. 인공지능망과 의사결정나무는 최선의 모형 선택을 자동화해주므로 사용의 용이성 측면에서 매우 우수한 방법이다.

16) 현재 인공지능망이 탑재된 통계 프로그램으로는 SAS, SPSS, R, Matlab, JMP 등이 있다.

제3장

국내외 선행연구 고찰

1. 회귀모형과의 비교 연구

1) 연구의 주요내용 및 특징

Lam, W. H. K. and Pan L. P. and Eddie C. M(2004)¹⁷⁾은 신경망모형을 이용하여 홍콩의 품목별 물동량을 예측하였고 이를 기존의 회귀분석에 의한 결과와 예측의 적합도 및 신뢰성 등을 비교·분석 하였다.

컨테이너 처리 물동량에서 싱가포르 항을 제치고 세계 최고를 기록하고 있는 컨테이너항만인 홍콩은 수십 년 넘게 항만개발계획 수립을 위한 항만물동량 예측에 있어 회귀분석모형을 주로 이용하여 왔다. 그러나 동 논문은 최근의 물동량 추세를 감안할 때 신경망모형이 우수할 것이라는 가설 하에 2002~2011년 기간을 대상으로 37개 품목에 대한 홍콩의 항만물동량을 신경망모형을 이용하여 예측하였다. 예측을 위해 1983~2000년 동안의 18개 시계열 자료를 이용하였다.

신경망모형을 이용해 예측된 항만물동량은 그 적합도의 검증을 위해 2001년의 실측치를 이용해 회귀분석 결과와 비교하였다. 몬테카를로 시뮬레이션을 이용하여 설명변수의 예측 오류 위험성에 따른 모형의 신뢰성(reliability)이 검토되었고, 세 가지의 성장률 시나리오를 설정하여 회귀분석에 의한 예측 결과

17) Lam, W. H. K. and Pan L. P. and Eddie C. M, "Forecasts and Reliability Analysis of Port Cargo throughput in Hong Kong", *Journal of Urban Planning and Development*, Sep. 2004.

와 비교하였다.

신성환(1995)¹⁸⁾은 최소자승예측법(OLS)을 통한 선형모형, 인공신경망모형(Artificial Neural Network), 그리고 이동평균법의 세 가지 방법에 의한 환율예측력을 비교·분석하였다. 추가적으로 신경망모형 추정에 쓰이는 두 가지 방법, 즉 비선형최소자승법(NLS)과 Back-Propagation(BP) 방법을 몬테카를로 실험을 통해 비교하였다.

인공신경망모형과 단순이동평균법의 두 가지 비선형예측기법을 사용하여 환율을 예측한 결과 인공신경망모형과 이동평균법 둘 다 수익률 예측 측면에서 효과적인 것으로 나타났으며 특히 인공신경망모형이 주별 예측에 유용한 것으로 나타났다.

신경망모형의 비선형최소자승법과 BP 방법을 몬테카를로 실험을 통해 비교·분석한 결과 유한표본에서는 비선형최소자승법이 BP 방법보다 수렴속도 및 R^2 측면에서 더 효율적인 것으로 나타났다.

김태훈·홍한국(2004)¹⁹⁾은 다양한 아파트 특성들에 대해 신경망모형을 이용하여 아파트 가격을 예측하고 기존의 회귀모형과 그 결과를 비교함으로써 회귀모형과 신경망모형의 상호연관성을 분석하였다. 이 논문은 아파트 가격을 예측하기 위하여 회귀모형과 신경망모형을 사용하였는데 두 모형의 정확도 면에서 신경망모형이 회귀모형에 비하여 전체적으로 더 적합한 것으로 나타났다. 그러나 아파트 가격을 예측하는 데 있어서 회귀모형과 신경망모형을 상호 보완적으로 사용할 필요성을 강조하였는데, 신경망모형이 예측의 정확성은 높지만 해석이 어렵기 때문에 정책적 판단 등을 내리기 어려운 문제점을 지적하였다.

손민우·이길성(2003)²⁰⁾은 신경망 이론과 다중회귀분석을 이용하여 일정

18) 신성환, “인공신경망모형과 이동평균법의 환율예측력 평가”, 「금융연구」, 제9권 1호, 1995.

19) 김태훈·홍한국, “회귀모형과 신경망모형을 이용한 아파트 가격 모형에 관한 연구”, 「국토연구」, 제43권, 2004, 12.

20) 손민우·이길성, “신경망 이론과 회귀분석을 통한 홍수위 예측”, 「대학토목공학회논문집」, 제23권 제3B호, 2003, 5.

지역을 대상으로 홍수위의 예측을 수행하며 각 예측방법별 예측 결과의 제공
 근 평균제곱오차(Root Mean Squared Error, RMSE), 평균절대오차(Mean
 Absolute Error, MAE), 편의(Bias) 등의 통계량을 통하여 예측의 적합성을 비
 교하였다.

홍수위의 수위 예측을 수행하기 위해 신경망 알고리즘 중 지도학습 과정이
 필요한 Back-Propagation 알고리즘, Flectcher-Reeves 알고리즘, BFGS 알고리
 즘을 이용하였다. 신경망모형을 통해 얻어진 결과는 실제값과 3.65%의 오차를
 보이며, 만족할 만한 결과를 보였다. 그리고 Stepwise 방식의 다중 회귀분석을
 통한 홍수위의 예측 결과는 실제값과 4.36%의 오차를 보였다.

신경망 알고리즘(Flectcher-Reeves 알고리즘)을 이용한 예측 결과가 다중회
 귀분석(Stepwise) 방식에 의한 예측 결과보다 예측정확도 면에서 우수한 것으
 로 나타났으나 그 차이는 그리 높지 않게 나타났다.

〈표 3-1〉 실제값에 예측값의 비율(161cm 기준)

단위 : %

예측방법	Ratio of RMSE(%)	Ratio of MAE(%)
신경망(Backpropagation)	11.58	6.06
신경망(Fletcher-Reeves)	4.70	3.65
신경망(BFGS)	5.79	4.55
다중회귀분석(Stepwise)	6.59	4.36

자료 : 손민우 외, “신경망 이론과 회귀분석을 통한 홍수위 예측”, 「대학토목공학회논문집」, 제23권 제
 3B호, 2003. 5.

2) 시사점

항만물동량을 예측하는 데 있어서 신경망이론과 회귀분석을 비교·분석한 결
 과 신경망이론을 통한 물동량 예측이 회귀분석 결과보다 신뢰성 및 정확성 면
 에서 전반적으로 나은 것으로 나타났다. 이것은 항만물동량이 각종 설명변수

들과의 관계에서 선형보다는 비선형의 관계에 가까운 것이 큰 영향을 미쳤을 것으로 판단된다. 손민우·이길성(2003)의 논문에서 보듯 수위에 대한 예측시 강우와 침투, 토양 특성, 상류에서의 수위 등과 같은 많은 매개변수들이 비선형으로 홍수위에 영향을 줄 경우 이를 해결하는 데 큰 효과를 볼 수 있음을 증명한 바 있다.

또한 인공신경망모형을 이용한 예측은 상대적으로 짧은 기간, 등락폭이 큰 기간에서 보다 예측력이 높은 것으로 나타났다. 따라서 항만물동량을 예측함에 있어서 신경망이론을 이용할 경우 비교적 짧은 간격에 등락폭이 큰 품목의 물동량 예측에 적용한다면 회귀분석 및 시계열분석의 단점을 더 효율적으로 보완할 수 있음을 시사하고 있다.

그러나 신경망이론의 예측력은 일반적으로 매우 높은 것으로 나타났지만 해석상의 어려움이 존재한다. 항만물동량을 예측하는 데 있어서 정확성 못지않게 변수의 상호 영향력 및 정책 수행의 효과를 분석할 필요가 있기 때문이다. 이러한 측면에서 필자들은 예측력을 강화시키기 위한 전략으로 신경망이론과 회귀분석을 상호보완적으로 수행할 것을 제안하고 있다.

2. 시계열모형과의 비교 연구

1) 연구의 주요내용 및 특징

노태협·이택호·한인구(2005)²¹⁾는 KOSPI 200 주가지수의 변동 예측을 금융시계열 모형과 인공신경망의 통합모형을 통해 제시하였고 이를 기존 금융시계열 예측방법론과 비교하였다. 주가지수의 변동 방향의 예측 면에서 금융시계열의 GARCH 모형이 인공신경망모형보다 우수한 성과를 보여 주었으나, 반

21) 노태협·이택호·한인구, “인공신경망-금융시계열 모형을 이용한 KOSPI 200 주가지수의 변동성 예측”, 「경영학연구」, 제34권 제3호, 2005. 6.

면에 인공신경망모형은 변동 예측의 정확성 면에서 GARCH 모형보다 월등히 앞선 것으로 나타났다.

〈표 3-2〉 모형간 예측력 비교

예측력	인공신경망(NN)	시계열모형(GARCH)	통합모형(NN-GARCH)
방향성(Hit Ratio)	41%	48%	56%
정확도(MAE)	0.33	0.27	0.33

자료 : 노태협 외, “인공신경망-금융시계열 모형을 이용한 KOSPI 200 주가지수의 변동성 예측”, 「경영학 연구」, 제34권 제3호, 2005. 6.

또한 이 논문에서는 인공신경망모형이 다양한 금융시계열 모형과의 통합을 통하여 변동 방향성 및 예측정확성의 동시적 추구 가능성을 제시하고 있다. 즉, 통합모형을 통해 인공신경망-금융시계열 통합모형은 모형의 예측력을 향상시킬 수 있고, 또한 기존의 단일인공신경망의 장점인 시행착오를 통한 최적 예측값 추정의 반복적인 과정(repetitive trial error)을 대폭 줄일 수 있다는 것이다.

정효준·이홍근(2002)²²⁾은 BOD를 예측하는 데 있어서 시계열모형(ARIMA)과 인공신경망모형의 예측의 적정성을 비교하였다. 예측의 적정성은 실제값과의 적합성뿐만 아니라 변동의 형태에 따른 적정성으로 판단하였다. 동 논문에서는 비선형모형인 ARIMA 모형과 인공신경망모형을 이용하여 청주 댐의 BOD에 근거하여 수질을 예측하였다. 1991년부터 2000년까지 월별 수질 자료를 이용하였으며 BOD값은 다층의 Back-Propagation 알고리즘을 이용한 신경망모형을 통해 예측하였다. 인공신경망모형은 두 개의 은닉 층으로 구성되었고 각 은닉 층은 15개의 노드로 구성하였다. 예측 결과 ARIMA 모형은 전반적으로 평균을 중심으로 변동하는 기간에 보다 적절하였고, 인공신경망모형은 최대와 최소값을 나타내는 기간에 보다 잘 부합되었다.

22) 정효준·이홍근, “ARIMA 모형과 인공신경망모형의 BOD예측력 비교”, 「한국환경위생학회지」, 제28권 제3호, 2002.

2) 시사점

상기 논문들은 기존의 예측방법론과 인공지능망이론과의 통합가능성을 보여주고 있다. 인공지능망모형의 설정을 위한 설명변수를 도출함에 있어서 시계열모형을 바탕으로 한 통계적인 분석을 이용하여 인공지능망모형과 시계열모형의 장점을 결합할 수 있다는 것이다. 즉 회귀분석의 경우와 같이 시계열모형에서도 신경망모형과 적절히 결합함으로써 예측의 정확도와 신뢰도를 높일 수 있다는 점이 강조되고 있다.

특징적인 것은 예측을 함에 있어 평균을 중심으로 변동하는 기간에는 전반적으로 시계열모형이 우수한 것으로 나타났고, 최대값과 최소값을 예측하는 능력은 인공지능망모형이 보다 나은 결과를 보여 주었다. 항만물동량은 품목에 따라 평균 또는 추세선을 기준으로 상하 변동하기도 하지만, 일부에서는 짧은 기간 내에 급등과 급락을 반복하는 형태를 보이기도 한다. 후자의 경우 회귀분석 또는 시계열 모형 보다는 인공지능망이 예측방법으로 더 적절할 것으로 판단하고 있다.

3. 기타 관련 연구

김경태·이장형·최현호(2001)²³⁾는 주가수익률의 예측을 통해 투자전략에 사용될 수 있는 인공지능망모형을 소개하고 모형의 성과를 검증하였다. 선행연구를 통해 각 기업의 당기순이익/자기자본 비율 등 58개 변수를 선정하였으며 1983~1996년간 14개 시계열자료를 분석에 이용하였다. 기간을 11개로 구분하되 시기별로 58개 변수 중에서 각 기업의 주가와 통계적인 유의성(t-test, 상관분석)을 검토하여 4~7개 변수를 선택하였다. 예측 결과 정확도의 범위는 기간

23) 김경태·이장형·최현호, “신경망에 의한 주가수익률 예측모형”, 「기업경영연구」, 제14집, 2001. 5.

에 따라 60~71% 사이 값을 기록하였고 평균은 64.96%로 나타났다.

해양수산부²⁴⁾는 국내외 경제여건 및 해운·항만환경의 변화를 감안하여, 컨테이너화물을 비롯한 전국 항만물동량의 예측과 컨테이너부두 등 전용부두의 적정하역능력을 재산정하였다. 이를 토대로 타당한 개발규모를 재검토하여 동북아 물류중심시대를 대비한 실천적 항만개발 정책방향을 수립하였다.

주요 연구내용은 우리나라 중장기 항만물동량 전망, 컨테이너부두 등 전용부두의 표준하역능력 재산정, 전국 항만시설 장기 수급전망, 동북아 주요 항만간 경쟁관계 분석, 동북아 물류중심시대의 신 항만개발 패러다임 제시 등이다.

전찬영(1999)²⁵⁾은 중장기 항만물동량을 거시적 방법에 의거, 항목별(수출, 수입, 연안)로 예측함으로써 경제상황 변화가 중장기 항만수요에 미칠 영향을 구체적으로 판단하였다. 각 항목별 물동량 특성을 고려하여 예측모형과 거시변수들 간의 다양한 조합을 통해 많은 모형을 설정 및 추정하였으며 여러 통계적 특성치들이 유의한 것으로 모형을 선정하였다. 또한 수출입 금액과 단위수출입 금액당 물동량 크기와의 함수관계를 규명함으로써 기 채택된 거시모형에 대한 교차 검증을 시행하였다.

1) 시사점

김경태·이장형·최현호(2001)는 신경망모형을 이용하기 위한 설명변수를 선정할 때 선행연구를 통해 다수의 변수를 후보군으로 우선 설정한 뒤 그 중에서 예측변수와 설명변수와의 통계적 유의성을 고려하여 최종 설명변수를 선택하는 방법을 택하였다. 후보군으로 설정된 변수는 이미 선행연구 과정을 통해 그 타당성이 충분히 인정된 것을 사용하였다.

전국 항만물동량 예측에서도 우리나라 32개 품목의 항만물동량을 예측할 때 품목별로 설정된 후보 모형 및 변수군은 선행 연구를 통해 타당성이 이미 널

24) 해양수산부, 「전국 항만물동량 예측」, 2004. 12.

25) 전찬영, “교차검증을 통한 우리나라 중장기 항만물동량 예측”, 「해양정책연구」, 제14권, 1999. 8.

리 알려진 것들이다.

전찬영(1999)은 국내총생산, 제조업생산액, 총고정자본 형성액, 광공업생산액, 실질 제조업 부가가치액, 대미달러 환율 등의 설명변수를 항만물동량의 주요 설명변수로 설정하였다. 예측모형에 사용된 함수형태는 회귀계수변동모형(time-varying coefficients regression model)을 비롯해 로그함수, 단순 회귀모형 등을 제시하고 있다. 본 연구에서는 이러한 예측모형을 일반적인 회귀모형을 변형시킨 일종의 변형된 회귀모형으로 간주하고 회귀모형과 신경망모형과의 예측 결과 비교와는 별도로 신경망모형과 비교·분석을 수행하였다.

4. 선행연구와의 차이점

신경망모형에 관한 선행연구는 대부분 선형모형 혹은 비선형모형(주로 시계열모형)과의 예측의 적합도를 비교하는 데에 주안점을 두고 있다. 특히 Lam, W. H. K. and Pan L. P. and Eddie C. M(2004)의 연구는 예측의 비교대상이 항만물동량인 점과 기존 물동량 예측모형인 회귀모형과의 비교 우위의 판단 기준을 예측의 적합도와 안정성(Reliability)에 두고 있다는 점에서 본 연구와 유사한 측면이 있다. 기존의 연구를 통해 신경망모형에 관한 예측의 정확도에 있어 비교우위를 어느 정도 인정받고 있으며, 분야에 따라서는 단기간 내에 심한 변동을 보이는 자료에 유효하게 이용될 수 있을 것으로 결론 내리고 있다.

그러나 대부분의 연구가 예측의 적합도만을 비교우위의 잣대로 삼고 있고, Lam, W. H. K. and Pan L. P. and Eddie C. M(2004)의 논문도 시뮬레이션을 통해 안정성을 함께 고려하고 있으나 예측의 민감도를 측정하는 강건성(Robustness)에 대한 고려는 하지 않고 있다. 신경망모형의 예측 우위를 평가하면서도 그 이유를 단지 모형의 비선형적 특성으로만 짐작하고 있을 뿐이다. 신경망모형의 최대 단점인 함수모형 부재에 따른 해석상의 어려움을 해결하는 방안으로 통합사용을 제시하고 있으나 구체적인 적용방안에 관한 내용은 언급

되지 않고 있다. 항만물동량의 예측방법으로 기존의 방법에 비해 예측력이 뛰어난 것으로 평가하고 있으나 품목별로 어떠한 변화 특성을 가진 물동량의 예측에 적합한지에 대한 분석도 이루어진 바가 없다.

따라서 본 연구는 신경망모형의 실질적인 활용 가능성을 높이기 위해 기존의 예측모형과의 비교·평가를 함에 있어 품목별 물동량의 변화특성을 유형별로 분류하여 분석함으로써 기존 이론과의 부합여부 및 특징적인 현상을 함께 도출하려 하였다.

신경망모형과의 비교 대상을 단순히 회귀모형에 국한하지 않고, 실제 예측에 많이 활용되고 있는 변형회귀모형도 포함시켜 상호 비교결과를 검토함으로써, 변수 변환을 통한 변형회귀모형이 단순 회귀모형에 비해 어떠한 변화를 가져왔는지를 분석한 것도 기존 연구에서 볼 수 없었던 내용이다. 또한 모형 간 우열의 판단기준에 적합도 외에 시뮬레이션을 통한 신뢰도 및 강건성을 함께 포함시킴으로써 판단의 객관성을 높이려 하였다. 이러한 본론의 연구결과를 바탕으로 중장기적 예측에 초점을 맞추어 온 현행 예측모형의 패턴을 어떻게 개선시킬 것인지, 신경망모형을 실제 예측에 적용할 경우 그 결과에 대한 원인 관계의 해석을 어떻게 해결할지에 대한 구체적 활용방안을 모색하는 것도 선행연구와 차별된다 할 수 있을 것이다.

제4장

주요 물동량 추세 변화 분석 및 설명변수 설정

1. 주요 품목의 선정

본 연구의 목적이 데이터 마이닝 기법에 의한 수리 예측모델을 품목별 항만 물동량에 적용시켜 예측의 신뢰도 및 정확도를 비롯한 다양한 통계적 특성들을 기존 모형과 비교해 봄으로써 동 기법의 유용성을 실증적으로 평가하고자 함을 전술한 바 있다. 따라서 본 서에서는 실증분석에 사용될 주요 품목으로 수입과 수출에서 각각 10개씩을 선정하였다.

품목의 구분은 해양수산부의 「해양수산통계연보」 상에 분류된 32개 품목을 대상으로 하되 품목을 선정하는 기준은 금액 단위로 각 품목의 전체 수출입에서 차지하는 비중(2005년 기준)이 큰 순서로 상위 10개 품목씩을 선정하였다. 이 때 품목별 금액산출은 관세청의 HS 코드를 「해양수산통계연보」 상의 32개 품목 분류에 대응시켜 구한 값이다(〈부록〉 참고). 단, 하역과정에서 안벽시설이 필요 없는 ‘유류’와 물동량 규모가 작아 항만시설소요가 크게 필요하지 않은 ‘항공기/선박 및 그 부품’은 품목선정에서 제외하였다.

수입물동량의 경우 수입금액을 기준으로 ‘전기기기 및 그 부품’, ‘기계류 및 그 부품’, ‘화학공업생산물’, ‘철강 및 그 제품’, ‘비철금속 및 그 제품’, ‘기타’, ‘방직용 섬유 및 그 제품’, ‘플라스틱/고무 및 그 제품’, ‘기타 광석 및 생산물’, ‘유연탄’ 등 10개 품목을 실증분석 대상으로 설정하였다. 수출물동량도 동일한 기준으로 ‘전기기기 및 그 부품’, ‘기계류 및 그 부품’, ‘차량 및 그 부품’, ‘플라스틱/고무 및 그 제품’, ‘철강 및 그 제품’, ‘화학공업생산물’, ‘방직용 섬유 및

그 제품', '기타', '비철금속 및 그 제품', '조제식품/음료/주류'등 10개 품목이 선정되었다.

2005년 기준 선정된 수입 10개 품목이 전체 수입금액에서 차지하는 비중은 65.6%(<표 4-1>참조)이며, 수출 10개 품목은 전체 수출금액의 87.1%를 점하고 있다. 항만시설 소요 산정에서 제외되는 석유류 품목을 제외하면 10개 품목이 수입과 수출에서 차지하는 비중은 각각 78.4%와 92.1%로 높아진다. 물동량으로 환산할 경우 이들 10개씩의 품목이 석유류를 제외한 전체 물동량에서 차지하는 비중은 수입과 수출이 각각 60.3%와 91.1%이다. 수출품목과는 달리 수입에서 금액단위 기준과 제법 비중이 차이 나는 이유는 양곡 및 철광석 등의 벌크화물과 목재류와 같이 부피에 비해 값이 싼 품목이 제외되었기 때문이다.

〈표 4-1〉 품목별 수입금액 비율 변동추이

순번	품목	연도		
		1995	2000	2005
	합계	100.00%	100.00%	100.00%
19	전기기기 및 그 부품	14.28%	22.14%	18.64%
2	원유(역청유), 석유	7.99%	15.72%	16.29%
18	기계류 및 그 부품	22.03%	17.23%	15.62%
28	화학공업생산물	8.49%	7.33%	8.00%
23	철강 및 그 제품	5.39%	3.82%	6.36%
4	석유가스 및 기타 가스류	1.61%	3.34%	4.04%
17	비철금속 및 그 제품	4.42%	3.42%	3.95%
32	기타	5.80%	4.45%	3.85%
3	석유정제품	2.79%	3.05%	2.89%
31	방직용섬유 및 그 제품	4.72%	3.35%	2.65%
29	플라스틱/고무 및 그 제품	2.74%	2.20%	2.65%
12	기타광석 및 생산물	1.32%	1.27%	2.06%
9	유연탄	1.47%	1.27%	1.84%
20	차량 및 그 부품	1.76%	1.22%	1.65%
14	조제식품, 음료, 주류 등	1.42%	1.26%	1.18%
11	목재, 목탄, 코르크 등	2.55%	1.70%	1.05%
21	항공기/선박 및 그 부품	2.98%	0.70%	0.96%
26	기타 동/식물성 생산물	1.15%	0.96%	0.88%
22	고철	0.74%	0.69%	0.87%
15	어패류, 갑각류 등	0.51%	0.77%	0.78%
1	양곡	1.37%	0.94%	0.75%
16	철광석	0.43%	0.59%	0.74%
30	피혁류 및 그 제품	1.70%	0.95%	0.60%
24	육류	0.55%	0.70%	0.55%
10	원목	0.77%	0.35%	0.27%
5	동/식물성 유지류	0.31%	0.14%	0.23%
27	당류	0.41%	0.24%	0.21%
8	무연탄	0.03%	0.05%	0.16%
6	비료	0.12%	0.10%	0.16%
7	시멘트	0.08%	0.02%	0.06%
25	제분공업생산물	0.05%	0.03%	0.04%
13	모래	0.02%	0.02%	0.02%

주 : 1. 순번은 해양수산통계연보상의 통계수록 순서임.

2. 전체 32개 품목 중 우리나라 수입금액에서 차지하는 비중(2005년 기준)을 기준으로 선정함. 단, '원유(역청유), 석유', '석유가스 및 기타가스류', '석유정제품'은 항만개발소요에 포함되지 않는 유류화물로 본 연구에서는 제외함.

자료 : 한국무역협회 무역정보네트워크(<http://www.kita.net>).

〈표 4-2〉 품목별 수출금액 비율 변동추이

순번	품목	연도		
		1995	2000	2005
	합계	100.00%	100.00%	100.00%
19	전기기기 및 그 부품	30.45%	26.92%	28.34%
18	기계류 및 그 부품	11.04%	18.36%	17.77%
20	차량 및 그 부품	8.23%	9.05%	13.23%
29	플라스틱/고무 및 그제품	5.47%	5.39%	6.32%
21	항공기/선박 및 그부품	4.65%	5.15%	6.18%
23	철강 및 그 제품	6.41%	4.88%	6.00%
3	석유정제품	1.90%	5.28%	5.42%
28	화학공업생산물	3.49%	4.27%	5.22%
31	방직용섬유및그제품	14.24%	10.54%	4.82%
32	기타	7.10%	4.97%	2.73%
17	비철금속 및 그제품	1.78%	1.92%	2.04%
14	조제식품, 음료, 주류 등	0.82%	0.63%	0.61%
30	피혁류 및 그 제품	2.10%	1.14%	0.36%
15	어패류, 갑각류 등	0.92%	0.61%	0.29%
26	기타동/식물성생산물	0.49%	0.30%	0.19%
12	기타광석및생산물	0.10%	0.08%	0.13%
27	당류	0.19%	0.10%	0.07%
6	비료	0.20%	0.07%	0.07%
7	시멘트	0.11%	0.07%	0.07%
22	고철	0.02%	0.01%	0.06%
11	목재, 목탄, 코르크 등	0.13%	0.07%	0.03%
4	석유가스 및 기타가스류	0.04%	0.12%	0.03%
25	제분공업생산물	0.01%	0.01%	0.01%
5	동/식물성유지류	0.02%	0.01%	0.01%
24	육류	0.08%	0.05%	0.00%
13	모래	0.00%	0.01%	0.00%
8	무연탄	0.00%	0.00%	0.00%
2	원유(역청유), 석유	0.00%	0.00%	0.00%
1	양곡	0.00%	0.00%	0.00%
10	원목	0.00%	0.00%	0.00%
9	유연탄	0.00%	0.00%	0.00%
16	철광석	0.00%	0.00%	0.00%

주 : 1. 순번은 해양수산통계연보상의 통계수록 순서임.

2. 전체 32개 품목 중 우리나라 수출금액에서 차지하는 비중(2005년 기준)을 기준으로 선정함. 단, '석유정제품'은 항만개발소요에 포함되는 않는 유류화물이며 '항공기/선박 및 그 부품'은 항만물동량에서 차지하는 비중이 미미하여 본 연구에서는 제외함.

자료 : 한국무역협회 무역정보네트워크(<http://www.kita.net>).

2. 품목별 물동량의 추세적 변화 특징 분석 및 설명변수 설정

1) 품목별 항만물동량 추세 변화의 구분

최근 들어 전통 제조업 성장세의 둔화와 IT, 첨단 신소재 등 고부가 신산업의 부상, 서비스 산업의 비중 확대 등 산업구조조정이 급속하게 이루어지고 있고, 제조공장의 해외 이전 확대, 경제 자유구역 지정 등 물류여건의 변화가 가속화되고 있다. 또한 중국 등 주변국가의 경제성장 지속과 더불어 이들 지역에 다국적 물류기지가 크게 확산되고 있고, 주요 교역국 간 FTA 협상의 확대, 글로벌 선사 및 운영사의 경영전략의 변화 등 우리나라 항만물동량에 영향을 미칠 수 있는 여건이 급속히 변화되고 있다.

이와 같은 기술혁신과 산업구조조정, 제조공장의 해외 이전, 물류체계의 변화는 업종별 산업군에 영향을 미치게 되고 결국은 해당 산업의 물동량에 많은 변화를 초래하게 된다. 이 같은 현상은 '90년대 후반 금융위기를 넘기며 특히 두드러지게 나타나고 있으며 품목별 물동량의 변화 추세에도 많은 변화가 나타나고 있다.

본 소절에서는 각 품목별 물동량의 변화추세를 살펴보고 당 원에서 기 발간한 연구보고서²⁶⁾ 내용을 토대로 당시 품목별 물동량 예측을 위해 사용하였던 설명변수(explanatory variable)와 품목별 물동량과의 관계를 재검토하였다. 제3장에서 언급한 바와 같이 각 품목별 예측을 위한 변수를 선정할 때 선행연구를 통해 다수의 변수를 후보군으로 우선 설정한 뒤 그 중에서 예측변수와 설명변수와의 통계적 유의성을 고려하여 최종 설명변수를 선택하였다. 후보군으로 설정된 변수는 이미 선행연구 과정을 통해 그 타당성이 충분히 인정된 것으로 간주하였다.

26) 해양수산부, 「전국 항만물동량 예측」, 2004. 12.

본 연구에서 살펴본 바로는 각 품목별 물동량의 변화추세가 다음의 형태를 크게 벗어나지 않은 것으로 판단되었다. 우선 '80~90년대의 상승추세를 최근에도 꾸준히 이어가거나, 연도별로 편차가 심하게 나타났지만 5년 내지 10년 단위의 단계별로는 과거의 상승추세가 유효한 것으로 판단되는 품목들을 꼽을 수 있다. 최근에도 상승추세를 이어가고는 있으나 과거에 비해 증가율이 현저히 둔화되는 품목, 그리고 어느 특정 시점을 기준으로 상하 진동만을 거듭하되 물동량이 상하 추세를 띄지 않는 품목들이 있다. 본 서에서는 이러한 품목별 물동량의 추세변화를 감안해 크게 4가지 변동 유형을 설정하여 품목별로 구분하였으며, 추후 회귀분석방법과 신경망분석과의 예측 결과 비교시 이러한 변동 유형과 연관시켜 분석하였다. 이것은 회귀분석과 신경망분석의 예측력이 물동량의 변화추세의 형태에 크게 영향을 받을 것이라는 가정이 전제된 것이다.

- 상승 지속형 : 물동량의 추세가 전반적으로 지속적인 상승을 하는 형태
- 상승 둔화형 : 전반적으로 상승하나 연도별로 증가율이 차츰 낮아지는 형태
- 불규칙 상승형 : 구간(단기)별로 변동이 불규칙적이나 단계별 저점이 높아지는 형태
- 정체 또는 불규칙 변동형 : 정체 내지 평균값을 중심으로 상하 변동을 하거나 전반적으로 증가나 감소를 하지 않는 형태

2) 수입물동량

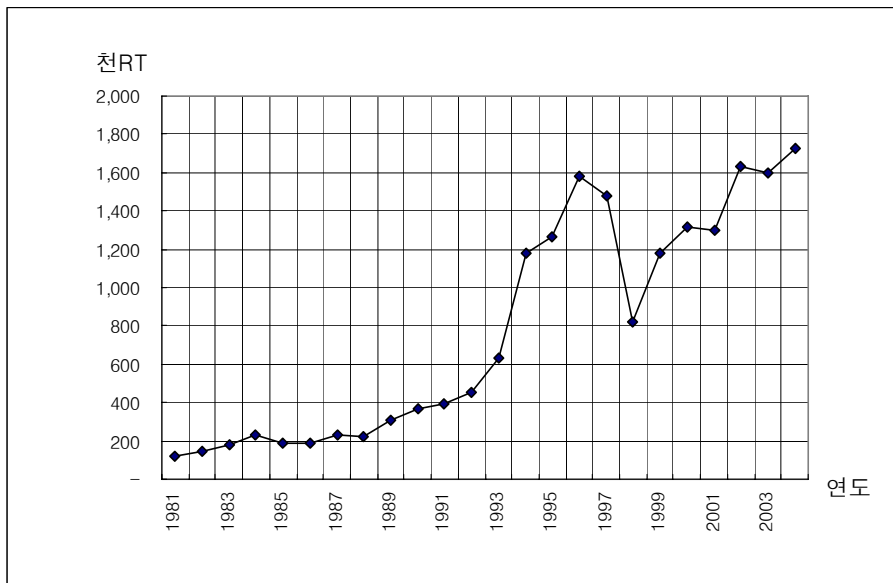
(1) 전기기기 및 그 부품(상승 둔화형)

전기기기 및 그 부품의 수입 항만물동량은 금융위기가 닥치기 직전인 1996년 150만 톤(R/T)으로 중기 고점을 기록한 후 다시 상승을 시작하여 약 6년이 지난 2002년에 이르러서야 전 고점을 갱신하였다. 경기회복세에 힘입어 2004년에는 전년 대비 7.8% 증가한 170만 톤으로 다시 사상 최고치를 기록하였으나 단기 고점이었던 '96년과 비교해서는 8년 동안 연평균 1.1%의 증가율을 기록함으로써 물동량 증가추세가 현저히 둔화되고 있음을 알 수 있다. 따라서 본 품목의 물동량 변화추세는 전형적인 '상승추세 둔화형'에 속하는 것으로 간주할 수 있다.

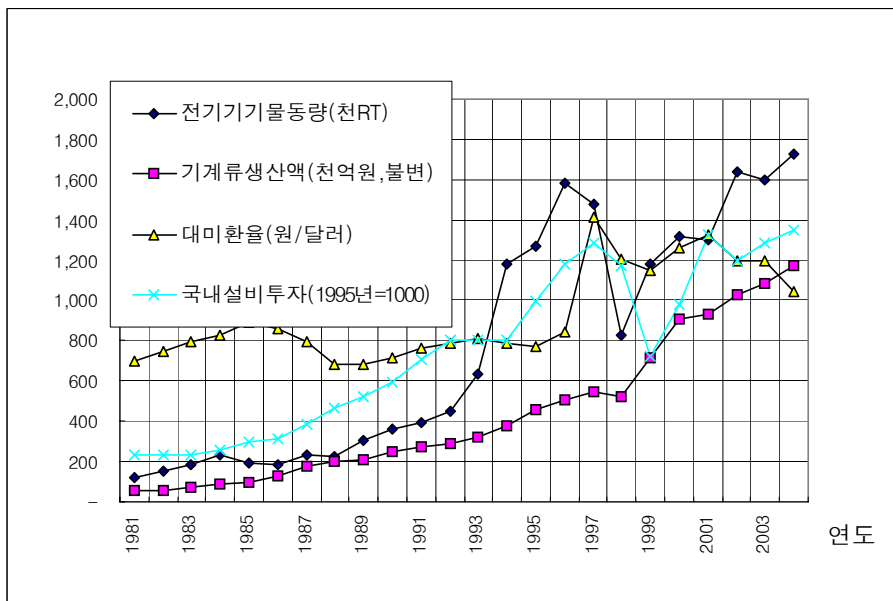
전기기기 및 그 부품의 수입은 대부분 중국산 저가 전기제품 및 전자부품이 주를 이루고 있는데 과거 전화기, 선풍기, 조명기기 등 단순 품목에서부터 최근에는 TV, 전동기, 발전기 등 비교적 고가 제품까지 가격경쟁력을 바탕으로 급속히 그 영역을 넓혀가고 있다. 전자부품의 경우 만성적인 대일·대미 무역적자가 지속되는 가운데, 국내 전자부품산업은 전반적으로 국산화율이 낮고 수입 의존도가 높으며, 노동집약적인 범용부품이 중국 등 해외투자 확대에 역수입이 증가하고 있다.

본 품목의 물동량 변화는 국내 설비투자의 흐름에 선행하되 상당히 유사한 움직임을 보이고 있다(〈그림 4-2〉 참조). 이는 주요부품의 국산화율이 낮고 제조원가의 상당부분이 원재료비이기 때문에 관련 완제품의 수출 및 내수의 증가에 따라 부품수입이 증가한 것과 유관한 것으로 판단할 수 있다.

〈그림 4-1〉 전기기기 수입물동량 변동추이



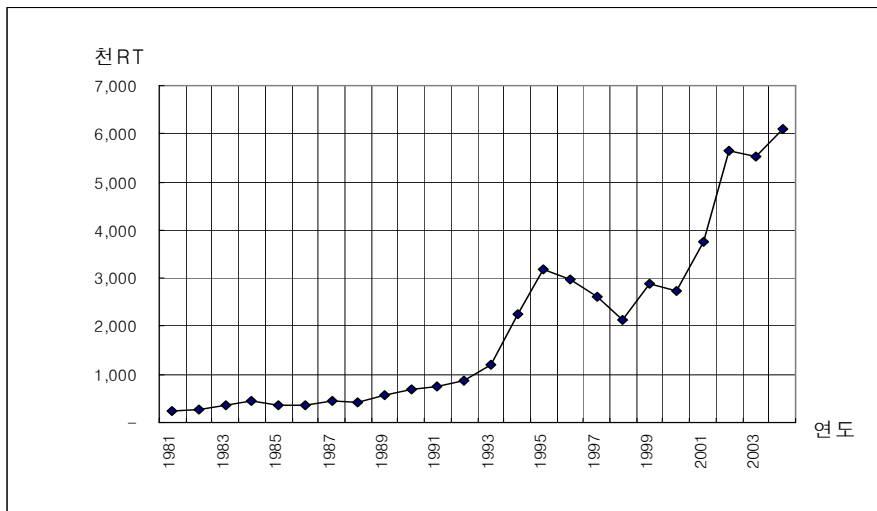
〈그림 4-2〉 전기기기 수입물동량 및 관련 변수 변동추이



(2) 기계류 및 그 부품(상승 지속형)

기계류 및 그 부품의 수입물동량은 금융위기를 전후한 1996~1998년을 제외 하고는 줄곧 상승하는 ‘상승 지속형’ 추세를 보이고 있다. 동 품목의 물동량은 1981년 이후 2004년까지 23년 만에 26배로 늘어났다. 주로 수입되는 기계류 품목은 일반 기계가 가장 규모가 크고, 그 다음으로 기관 및 터빈, 풍수력기계, 금속공작·가공기계 순이다. 수입되는 기계류는 산업전반에서 필요로 하는 생산 설비를 제조·공급하는 전형적인 자본재 산업으로서 산업경기의 선행지표가 되는 산업이며, 환경규제 및 SOC 투자 등 국가 정책사업과도 밀접하게 연계되는 산업이다. 따라서 기계류 물동량(혹은 생산액)과 국내 설비투자와의 변동추세는 상하 움직임의 폭에 있어서는 제법 차이를 보이고 있으나 흐름의 방향은 거의 동일한 것을 알 수 있다(〈그림 4-4〉 참조).

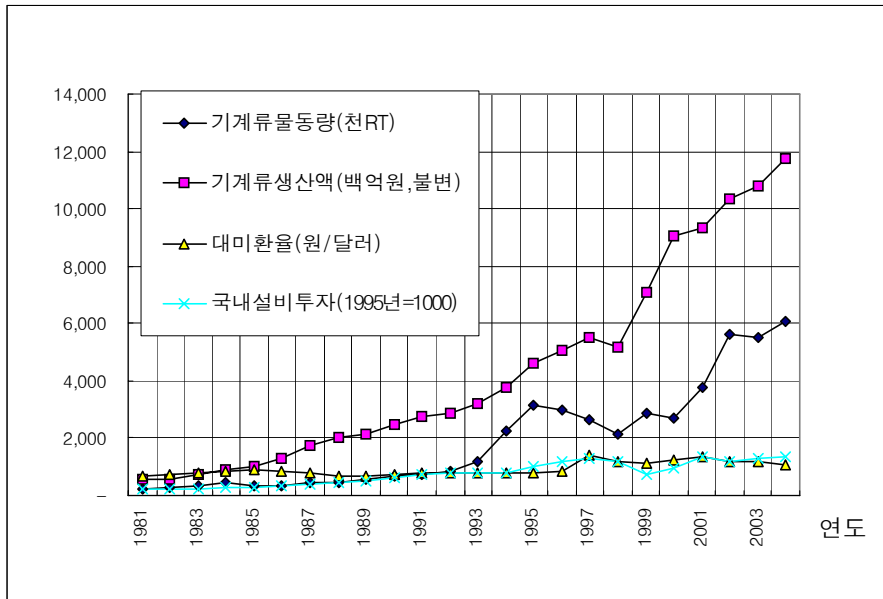
〈그림 4-3〉 기계류 수입물동량 변동추이



기계산업은 전후방 연관산업과 긴밀한 기술교류와 산업협력이 요구되는 산업으로 최근 들어서는 첨단기술과 복합화·융합화되어 다기능화와 시스템화를

추구하고 있다. 즉 전자·전기기술, 정보통신기술의 발달과 기계·자동차부품의 디지털화 등 관련 기술의 발달에 따른 부품·소재기술 간 융·복합화 현상이 증대되고 있다. 그러나 이러한 첨단·복합기능을 겸비한 기계류에 대한 소비가 증가할수록 60~70% 수준(기종에 따라서는 70~90%)에 머물러 있는 자립도 혹은 기술수준에 비추어 해당 품목에 대한 수입증가 현상은 지속될 것으로 전망된다.

〈그림 4-4〉 기계류 수입물동량 및 관련 변수 변동추이



(3) 화학공업생산물(상승 둔화형)

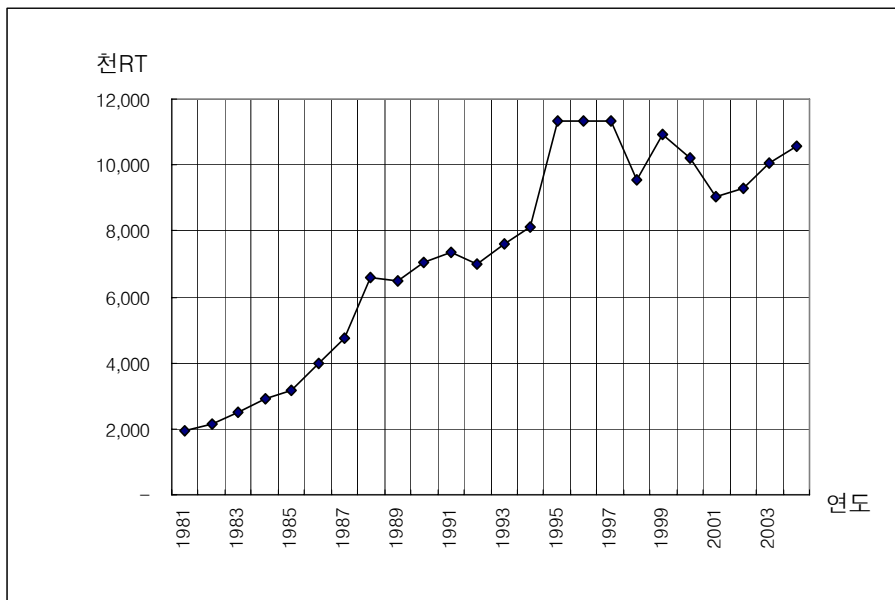
화학공업생산품의 수입물동량은 1995년 1,135만 톤으로 정점에 이른 후 900만~1,100만 톤 사이에서 움직이다가 2001년을 저점으로 다시 회복세를 나타내고 있다. 그러나 '95년에 기록한 고점에 아직 미치지 못하고 있어 전형적인 '상승 둔화형' 형태의 물동량 변화추이를 보이고 있다. 본 품목에서 화공생산품이란 무기, 유기화합물, 의약품, 염료, 페인트, 잉크, 향료, 비누, 카페인, 알부

민, 전분, 효소, 성냥, 필름, 인화지 등을 의미한다.²⁷⁾

2003년 금액기준으로 각 화학공업생산품이 차지하는 비중은 유기화합물(39%), 각종 공업생산품(18%), 무기/유기화합물(12%), 페인트/잉크(8%), 의료용품(8%) 순이다.

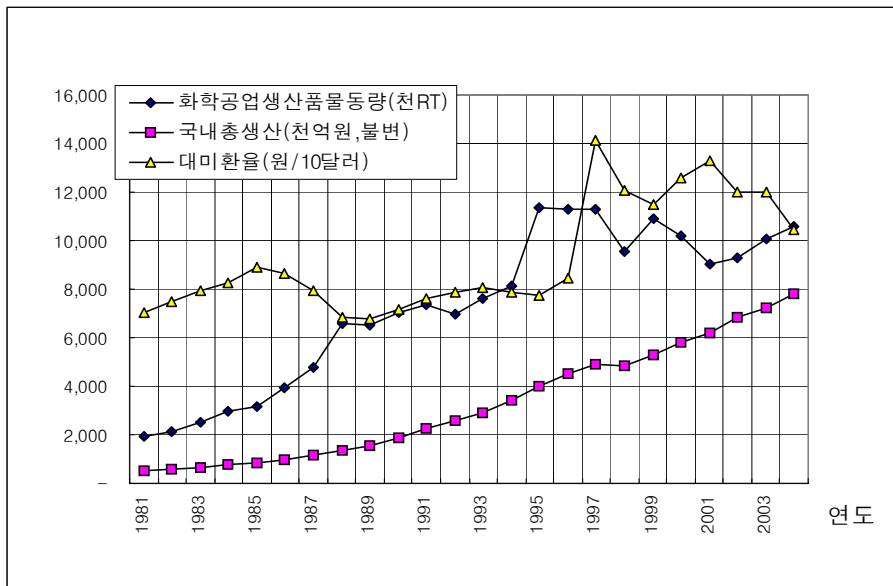
동 품목은 주로 국내에서 직접 소비되기도 하지만 상당량은 다시 원부자재로 투입되어 생산·가공된 후 최대 수출지인 중국 등지로 수출된다. 즉 국내에 수입되는 화학공업생산품의 세부품목은 중간재와 소비재 모두를 포함하고 있다. 따라서 동 품목의 수입물동량은 국내소득 증가율(GDP로 대변) 및 수출경쟁력과 관련해 환율의 변동에 상당한 영향을 받는 것으로 평가되고 있다.

〈그림 4-5〉 화학공업생산품 수입물동량 변동추이



27) 석유화학산업의 주요 생산품인 합성수지, 합성섬유, 합성고무는 화물수송통계상 '플라스틱, 고무 및 그 제품', '방직용 섬유 및 그 제품' 으로 집계됨.

〈그림 4-6〉 화학공업생산물품 수입물동량 및 관련 변수 변동추이



(4) 철강 및 그 제품(상승 지속형)

철강 수입물동량은 국내의 경기순환변동과 1980년대 후반 신도시 건설 및 사회간접자본 확대, '90년대 후반의 금융위기에 따른 경기침체 등 내부적인 요인에 따라 등락 현상을 반복해 왔으나 전반적으로 상승추세를 지속하고 있는 '상승 지속형' 변화추이를 나타내고 있다.

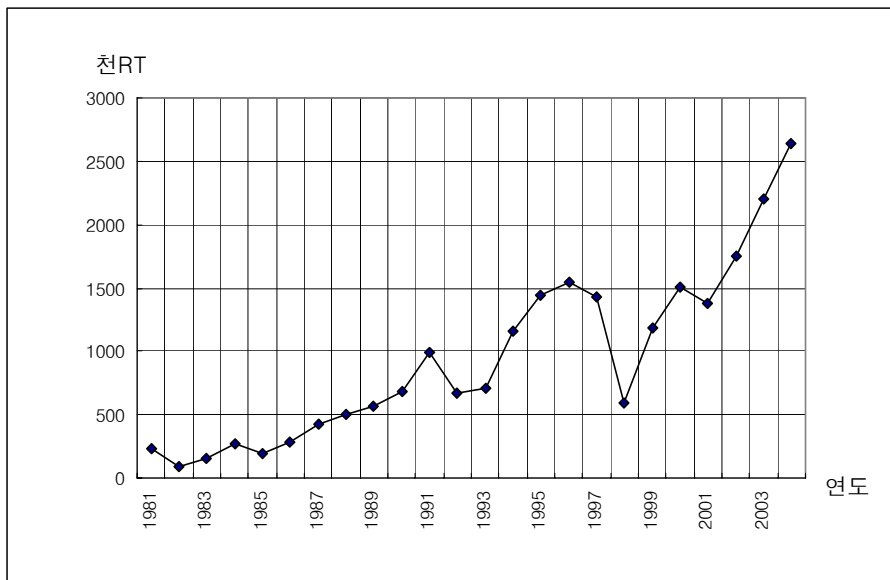
단계별로 살펴보면 '80년대 들어 중반까지는 국내 철강산업의 설비증설과 특수강 등 강종 개발에 따른 자급도 향상으로 물동량의 증가세가 미미하였다. 그러나 '87년 이후 경제적 호황기와 함께 신도시 건설, 사회간접자본 투자확대 등 건설경기의 상승과 고급강을 중심으로 한 특수강에 대한 수요 증대로 금융 위기가 닥치기 직전인 1996년까지 비교적 큰 폭의 증가추이를 보여 왔다.

2000년대 들어서는 건설, 자동차, 조선 등 수요산업의 생산이 호조를 보이며 철강수요가 꾸준히 증가하고 있는 데 반해, 철강생산이 이를 충분히 뒷받침하지 못하면서 핫코일 및 강판제품 위주로 수입물동량이 다시 급증하는 추세를

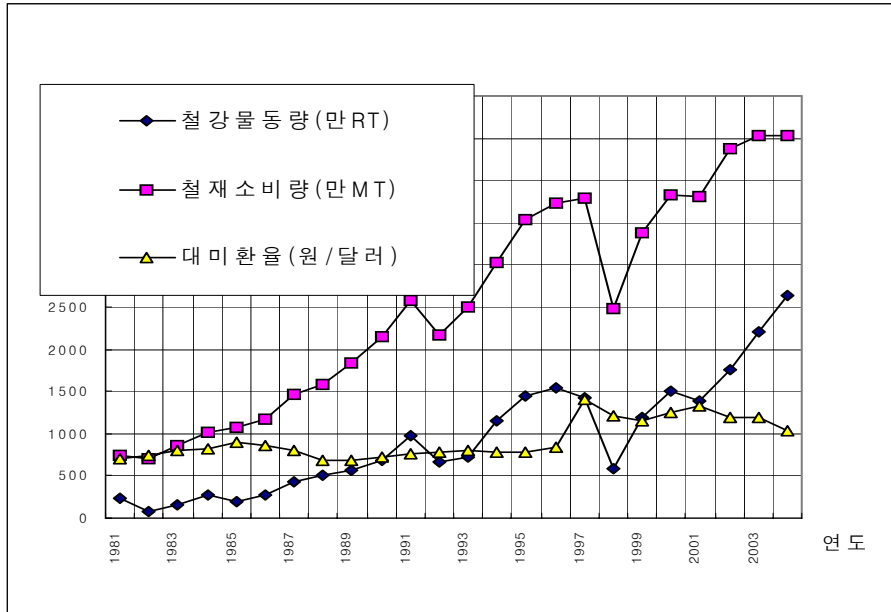
보였다.

우리나라는 국민 1인당 철강소비가 대만에 이어 세계 2위권 수준을 차지할 만큼 철강과 연계된 산업군(조선, 자동차 등)의 국민경제적 기여도가 높은 편이다. 특히 보통강 부문에 세계 최고의 경쟁력을 보유하고 있고 고효율의 설비 및 우수인력을 보유하고 있다. 반면 고급강 부문의 기술력이 여전히 취약하고, 생산제품의 고급화를 추진하고 있으나 신규설비 투자가 매우 저조한 편이어서 향후 철강수요가 지속될 경우 중국 등지로부터 저가 철강제품의 수입 증가세는 당분간 지속될 가능성이 높을 것으로 전망된다. <그림 4-8>에서도 알 수 있듯이 철강수입 물동량의 증가추이는 국내 철강수요와 거의 같은 흐름을 보이고 있다.

〈그림 4-7〉 철강 수입물동량 변동추이



〈그림 4-8〉 철강 수입물동량 및 관련 변수 변동추이

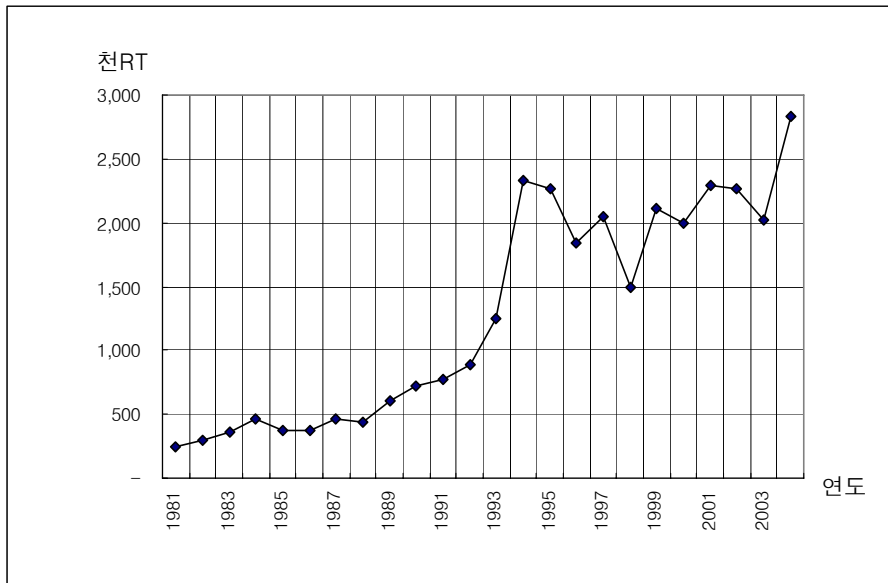


(5) 비철금속 및 그 제품(불규칙 상승형)

비철금속 산업은 자동차, 전기, 전자산업, 중화학공업 등의 필수 기초소재인 알루미늄, 전기동, 아연, 연, 기타금속 등을 공급하는 장치산업이며, 최근 들어서는 비철금속 첨단산업의 기초소재 공급원으로서 그 중요성이 더욱 높아지고 있다. 그러나 우리나라는 비철금속의 자급률이 낮아 해외수입에 의존하는 비중이 높을 수밖에 없다. 비철금속 생산이 미미하여 자급률이 낮았던 '90년대 초중반까지 자동차, 전기·전자산업 등 관련 산업의 지속적 발전으로 비철금속 수입물동량 역시 꾸준한 상승세를 이어왔다. 국내 비철금속 산업군이 꾸준히 성장하여 생산능력이 확충되고 자동차, 전자 등 관련 산업의 해외 생산의 확대와 함께 비철금속의 주요 수요처 이동현상이 함께 발생하면서 비철금속 수입물동량은 1994년을 정점으로 정체현상을 보이고 있다. 그러나 2004년 물동량은 전년 대비 큰 폭으로 늘어나면서 사상 최고치를 기록하였는데 이것은 관련

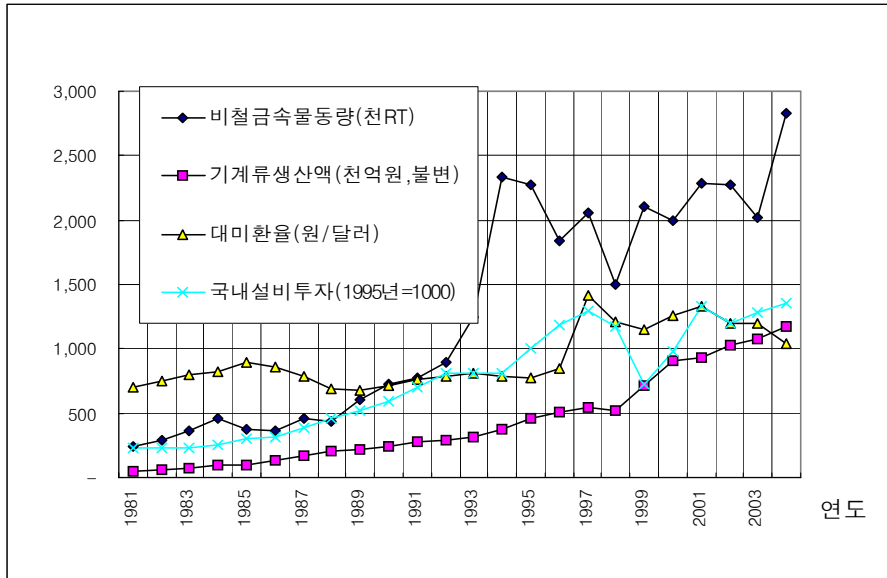
철단산업의 급속한 발전에 따른 수요 증대와 광양에 유치된 LME²⁸⁾ 지정창고의 영향 때문인 것으로 판단된다. 비철금속 물동량의 변화추이는 불규칙하나마 저점을 단계적으로 높여가는 ‘불규칙 상승’형에 가깝다. 비철금속 수입물동량은 관련 산업군인 기계류 산업군의 생산 증가율과 가장 밀접한 상관관계를 보이고 있다.

〈그림 4-9〉 비철금속 수입물동량 변동추이



28) LME는 동, 알루미늄, 아연, 니켈 등의 세계 비철금속 가격의 현물기준을 제시하여 실물 거래자들에게 거래단가를 제공해 주며 이들에게 선물거래를 허용함으로써 가격변동으로 발생하는 위험을 회피하기 위한 헤징의 기회를 제공하고 있다. LME 지정창고는 공급부족으로 실물시장이 위축될 때 산업에 대한 최후의 공급자 역할을 하는데, 불황의 경우 산업은 창고에 여유물량을 보관하여 은행이 동 물자의 금융비용을 부담하도록 하고 있다. 현재 미국의 LME 지정창고를 비롯해 13개국 43개 지역에 400여개 이상의 LME 지정창고가 운영되고 있다.

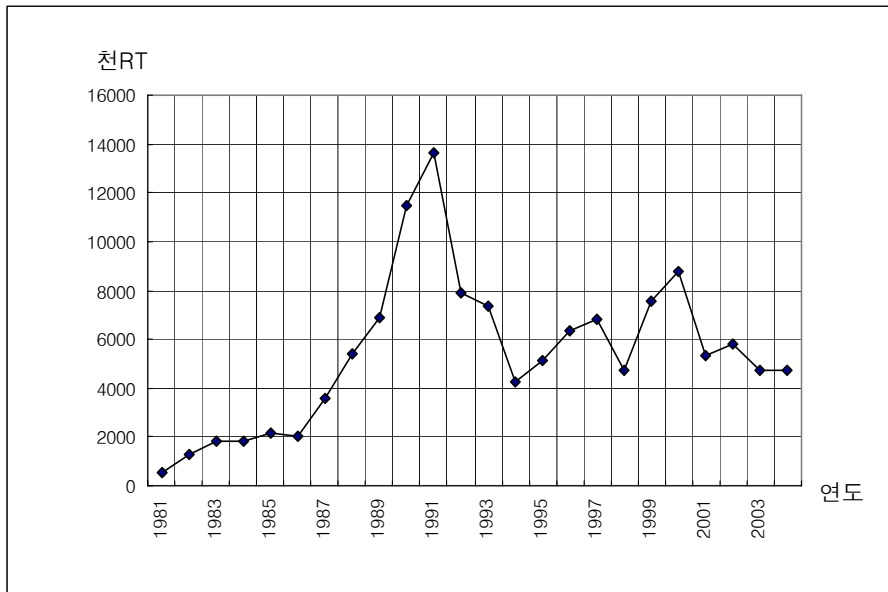
〈그림 4-10〉 비철금속 수입물동량 및 관련 변수 변동추이



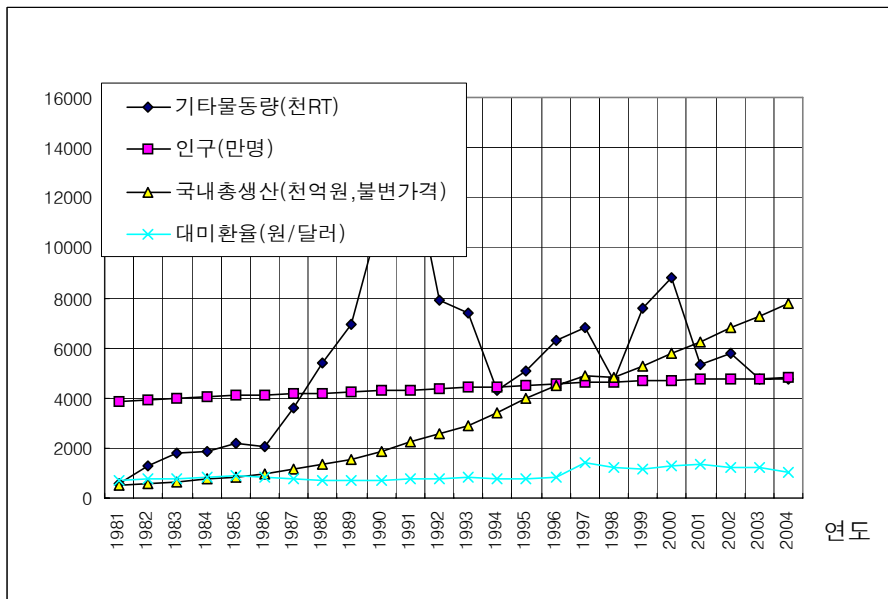
(6) 기타 잡화(정체형)

기타 잡화는 31개 품목 분류에서 제외된 여타 품목으로 펄프, 판지, 우산, 지팡이, 도자기, 유리, 석면, 석재, 악기, 완구, 예술품 등 다양하게 구성되어 있다. 따라서 여러 특성을 지닌 품목이 서로 복합되어 있기 때문에 특정 산업군에서 엿볼 수 있는 특징을 찾아내기 어렵다. '80년대 이후 '90년대 초까지 지속적인 상승 추세를 보이다 '91년 1,366만 톤(R/T)을 고점으로 크게 하락하기 시작한 후 '90년대 중반 이후 400만~800만 톤에서 등락을 반복하고 있다. 기타 잡화 수입물동량은 600만 톤의 평균치를 중심으로 특정 구간 내에서 상하 편차를 보이고 있는 '정체'형 행태의 물동량 추세를 나타내고 있으며 유사한 흐름을 보이고 있는 특정 경제지표를 설정하기가 어렵다.

〈그림 4-11〉 기타 수입물동량 변동추이



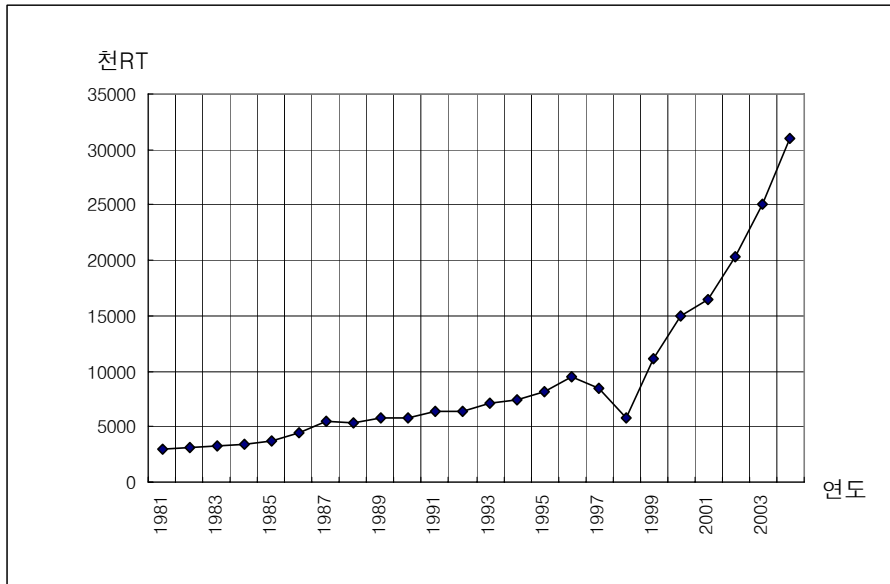
〈그림 4-12〉 기타 수입물동량 및 관련 변수 변동추이



(7) 방직용섬유 및 그 제품(상승 지속형)

동 제품은 외환위기를 겪었던 '97년 안팎을 제외하곤 줄곧 상승세를 이어오는 전형적 '상승 지속형' 물동량 추세를 보이고 있다. '80년대에 비해 외환위기 이후 더욱 가파른 물동량 증가세를 나타내고 있는데, '80년대 후반부터 해외로 생산기지를 이전한 업체 제품들의 역수입과 국민소득 수준의 향상과 수요의 다양화·고급화 추세에 따라 의류 등 고급제품의 소비증가에 기인한 바 크다. 또한 레저 스포츠웨어 및 산업용, 비의료용 수요가 급증하고 있고, '93년 하반기부터 제3차 유통시장 개방을 계기로 해외 브랜드 제품과 중국 등에서 저가의류의 유입이 급증하고 있다. 수출에 따른 섬유원료 등 원자재 수입 증가와 소비자들의 외국 유명 브랜드 선호로 인한 이태리, 프랑스 등의 고가 패션 의류 수입량도 계속해서 늘고 있다.

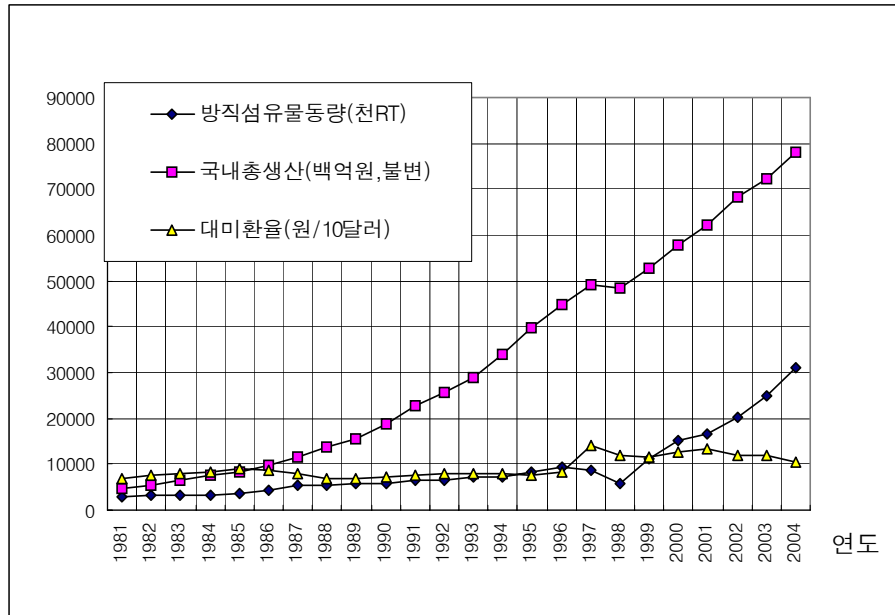
〈그림 4-13〉 방직용섬유 수입물동량 변동추이



향후 동 제품의 수입은 주요 교역국가와의 FTA의 확산, 국내시장의 개방과 선진국 대비 낮은 관세율(8%)에 따라 증가추세를 이어나갈 것으로 보인다. 동

제품은 대표적인 소비재로서 국민 소득수준의 척도인 GDP와 원화의 환율가치에 의해 많은 영향을 받는다.

〈그림 4-14〉 방직용섬유 수입물동량 및 관련 변수 변동추이



(8) 플라스틱/고무 및 그 제품(정체형)

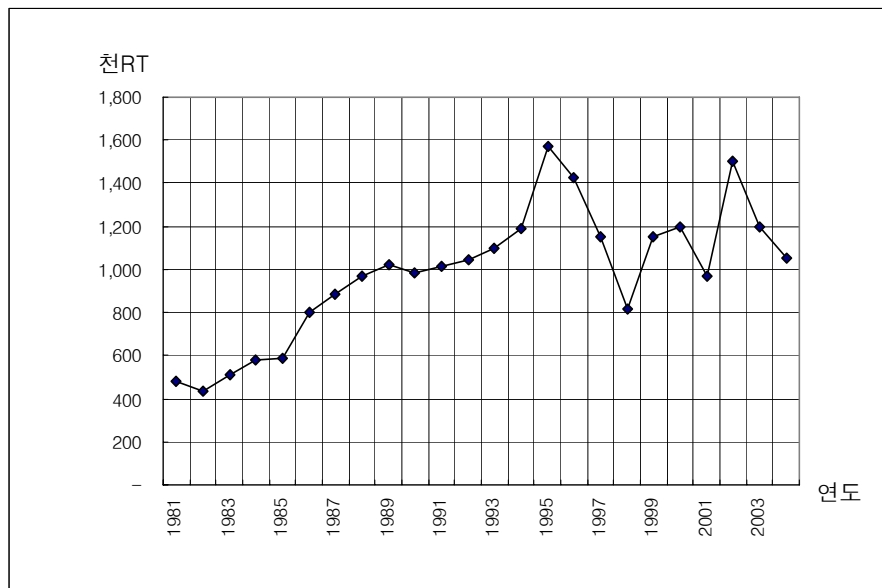
동 제품의 플라스틱류는 일차제품인 에틸렌의 중합체에서 플라스틱제 포장용기, 물품운반기기, 기타 가정용품에 이르기까지 매우 다양하다. 고무류는 천연고무, 합성고무, 재생고무를 비롯하여 고무제 위생용품, 의료용 등을 포괄하고 있다.

동 제품은 '80년대 이후 최고점을 기록한 1995년까지 거의 지속적인 물동량 상승세를 보여 왔다. 외환위기 이후 국내 경기의 급속한 회복세를 바탕으로 잠시 증가세로 돌아서기도 했으나 '95년의 고점을 갱신하지 못한 채 120만 톤을 중심으로 등락을 반복하고 있는 '정체형' 움직임을 보이고 있다. 그 원인으로

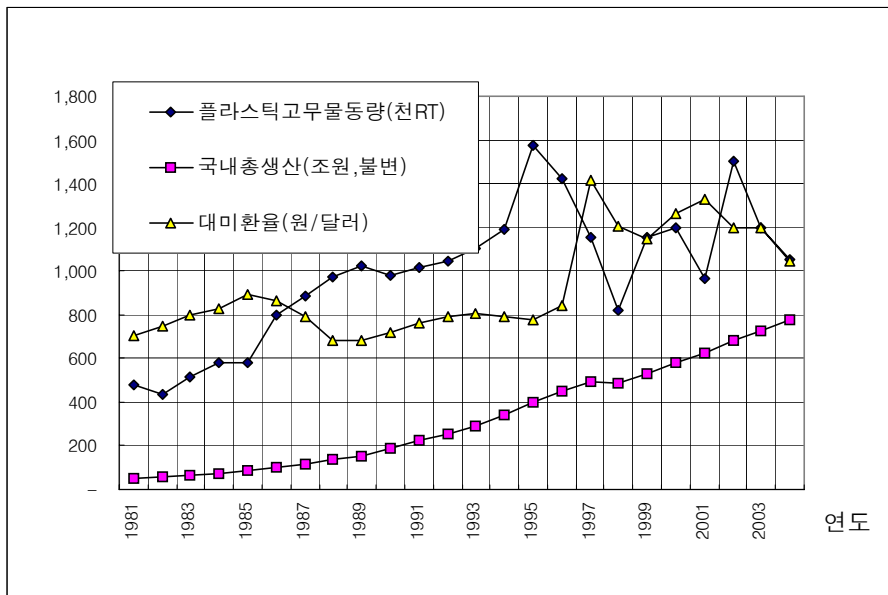
플라스틱 및 합성고무의 원재료가 되는 에틸렌의 국내 생산량이 증가하였고 국민의 소득수준 향상으로 인해 특정 품목을 제외하고는 플라스틱 제품의 국내수요가 둔화된 점이 지적되고 있다.

플라스틱/고무류는 직접 소비재로 이용될 뿐 아니라 각종 제조업 생산의 중간 원자재로 사용되고 있으므로 향후 수입물동량은 대내적으로 제조업 생산과 인구에, 대외적으로는 대미 달러 환율에 영향을 받을 가능성이 높다.

〈그림 4-15〉 플라스틱고무 수입물동량 변동추이



〈그림 4-16〉 플라스틱고무 수입물동량 및 관련 변수 변동추이



(9) 기타광석 및 그 제품(상승 지속형)

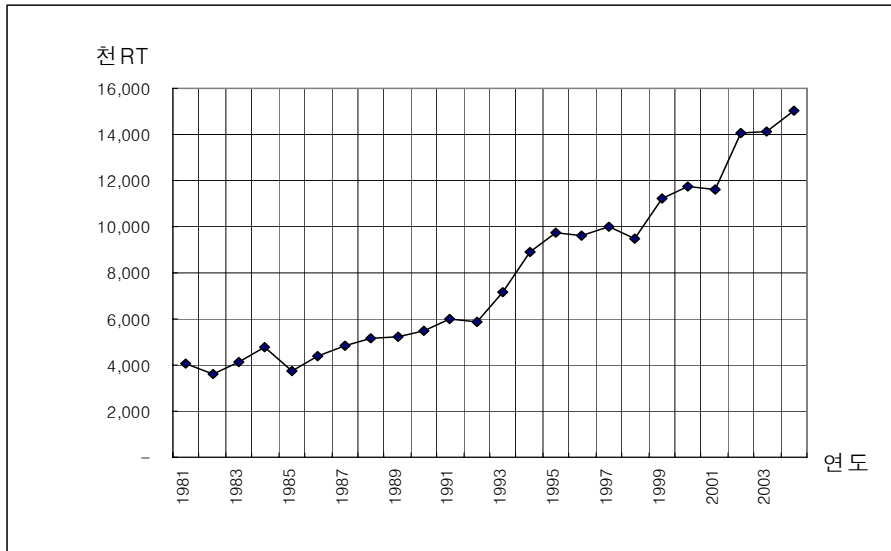
동 제품은 소금, 인광석, 기타광석을 의미한다, 소금은 수입물량의 대부분이 공업용 원료인 소다의 원료로 사용되고 약 10% 정도만 일반시판용 소금으로 이용되고 있다. 인광석은 주로 비료의 원료로 사용되고 있는데 그 자체가 비료로 쓰이는 것이 아니라 인산액으로 용해하여 복합비료를 제조하는 데 이용되고 있다. 소금, 인광석을 제외한 기타광석은 주로 아연광, 규사, 석고 등이다.

이 중 인광석은 경지면적의 용도 전환추세와 복합비료 소비가 유기질비료로 대체되는 상황이 반영되어 차츰 감소하는 것으로 전망되고 있다. 이에 반해 소금 및 물동량 비중이 가장 높은 기타광석(소금, 인광석 제외)에 대한 수요는 비철금속 산업 등 관련 산업의 발전으로 지속적인 상승세를 보이고 있다.

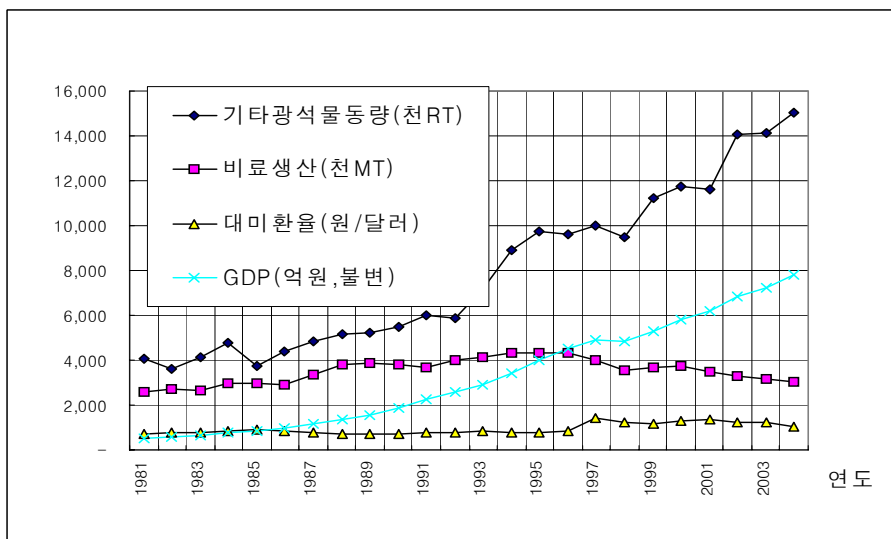
동 제품은 '80년대 이후 짧은 기간(2~3년)의 조정과 상승을 반복하는 '상승 지속형' 물동량 추이를 보이고 있다. 비료생산의 감소로 인광석에 대한 수요는

줄고 있으나 동광, 규사, 아연 등 첨단소재산업의 발달에 따른 관련 산업의 여파로 향후에도 지속적인 상승이 예상된다.

〈그림 4-17〉 기타광석 수입물동량 변동추이



〈그림 4-18〉 기타광석 수입물동량 및 관련 변수 변동추이

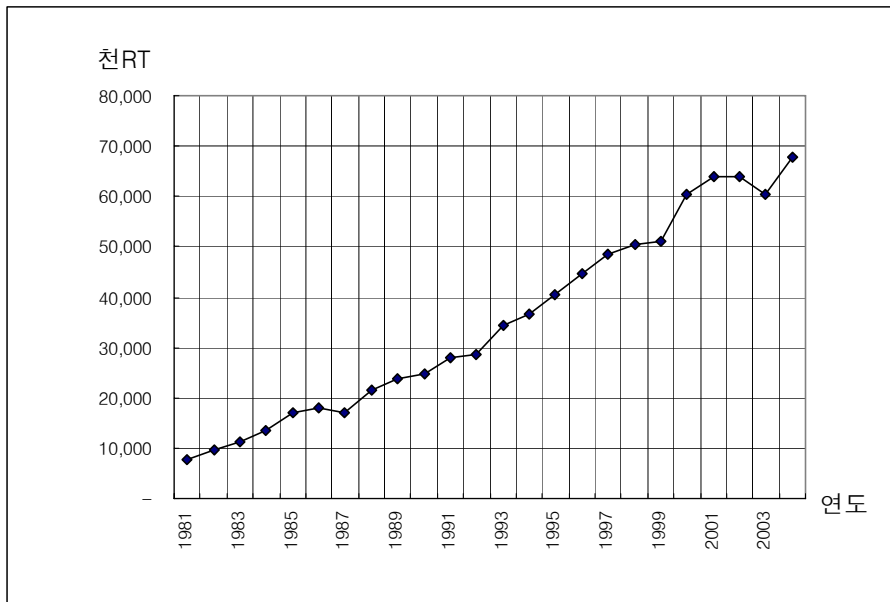


(10) 유연탄(상승 지속형)

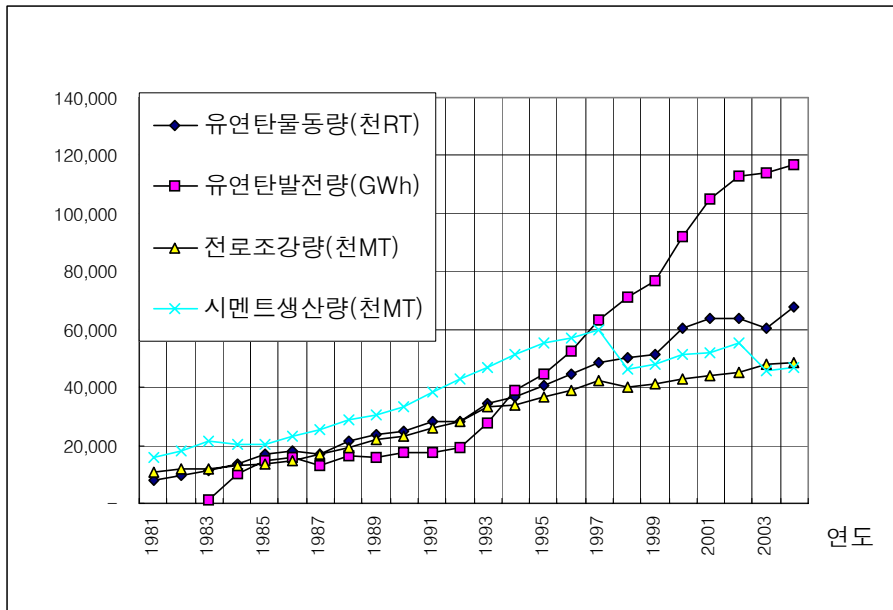
유연탄의 수요는 크게 발전용, 제철용, 시멘트 및 기타 생산용으로 구분되며 전량 수입에 의존하고 있다. 이 중 절반 이상의 비중을 차지하고 있는 발전용은 제 1·2차 석유파동으로 원유수입을 줄이기 위해 1983년 국내 최초의 유연탄 발전소인 삼천포 발전소가 건설된 이후 보령발전소와 호남발전소가 전력을 생산하면서 본격적으로 수요가 증가하였다. 약 30%의 비중을 점하고 있는 제철용과 그 나머진 시멘트 생산용 및 기타 용도의 유연탄 수요는 상승세가 점차 둔화되고 있다.

전체적으로는 제철용 및 시멘트용 수요증가의 둔화에도 불구하고 전력생산에서 유연탄의 발전량 비중을 높이려는 정부 정책에 힘입어 '80년대 이후에도 유연탄의 수입물동량은 지속적으로 증가세를 유지하는 '상승 지속형' 패턴의 물동량 추세를 보이고 있다.

〈그림 4-19〉 유연탄 수입물동량 변동추이



〈그림 4-20〉 유연탄 수입물동량 및 관련 변수 변동추이



3) 수출물동량

(1) 전기기기 및 그 부품(불규칙 상승형)

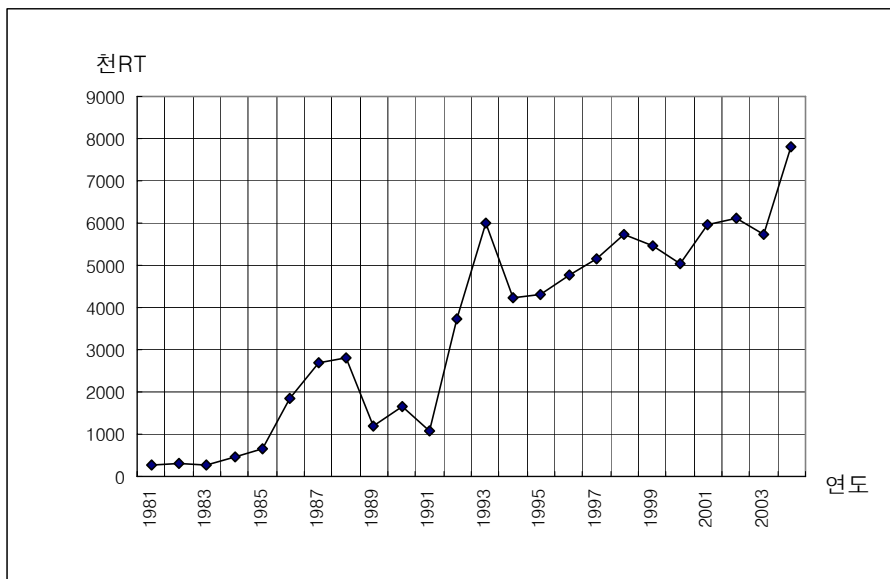
전기기기 제품보다는 전기 혹은 전자부품산업에 훨씬 강점을 보이고 우리나라는 세계적인 IT 산업의 활성화, 기술혁신과 정부의 적극적인 지원정책에 힘입어 국내 전자부품산업도 급격히 성장할 것으로 기대되고 있다. 중국과 인도, 그리고 유럽 등의 시장비중이 확대되고 있으며, 세계의 공장화 및 시장화의 흐름 속에서 특히, 중국의 비약화로 국내 관련 제품의 중국 진출이 가속화되고 있다.

'70년대 이후 수출 위주형 제조산업의 육성으로 전기·전자산업의 부흥기를 맞으면서 '90년대 초반까지 관련 제품의 수출물동량이 급증세를 보였으나 1993년 600만 톤(R/T)으로 고점을 이룬 후 거의 10년간을 횡보하는 양상을 보

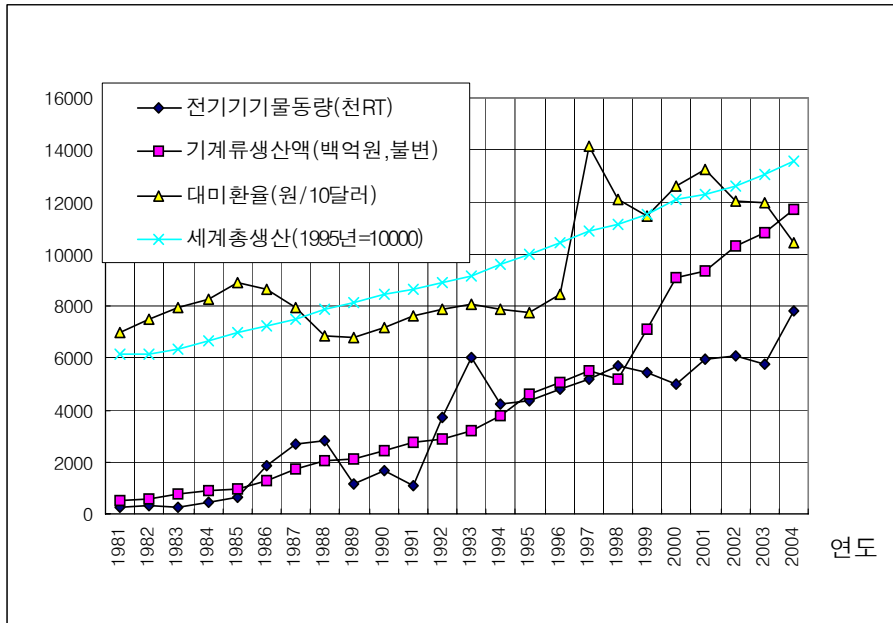
였다. 그러나 세계경기가 호황을 이루면서 LCD, 브라운관 등의 능동부품을 중심으로 시장이 확대되고 정보통신과 관련하여 수동부품의 수요가 확대되면서 동 제품의 수출물동량은 2004년 780만 톤으로 '93년의 고점을 갱신하였다.

'90년대 초반 이후 전반적으로 수출물동량의 증가세가 둔화되는 양상을 보이고 있으나 아직 상승추세가 유효한 '불규칙 상승형'의 흐름을 이어 나가고 있는 것으로 판단된다.

〈그림 4-21〉 전기기기 수출물동량 변동추이



〈그림 4-22〉 전기기기 수출물동량 및 관련 변수 변동추이



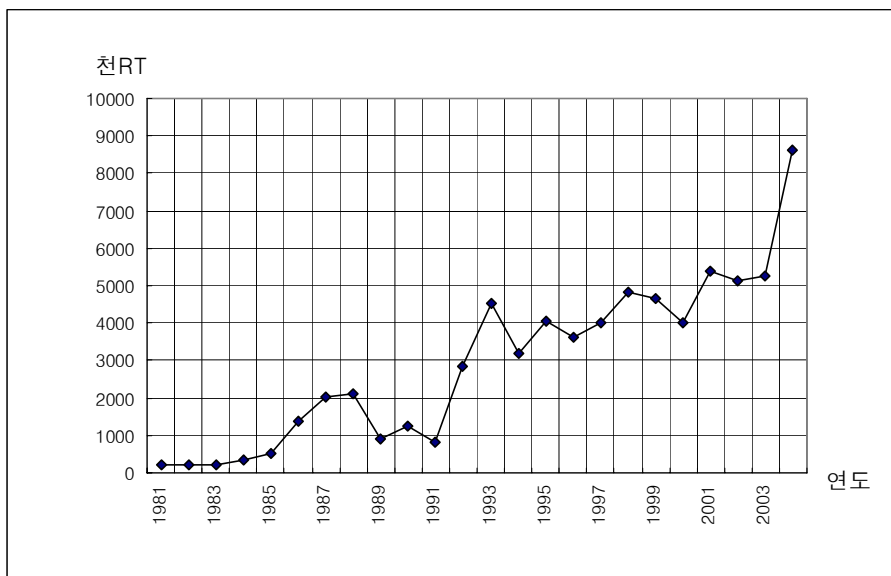
(2) 기계류 및 그 부품(불규칙 상승형)

기계류 산업은 산업 전반에서 필요로 하는 생산설비를 제조·공급하는 전형적인 자본재 산업이다. 구성품목 중 반도체 제조용 기계와 내연기관 및 터빈이 가장 높은 성장률을 기록하였고, 금속주조 및 기타 야금용 기계, 탭·밸브 및 유사장치, 기타 일반기계 순의 성장세를 보였다. 수출 면에선 냉동·공조기계, 건설·광산기계, 섬유기계, 풍수력기계 순으로 수출이 되고 있다.

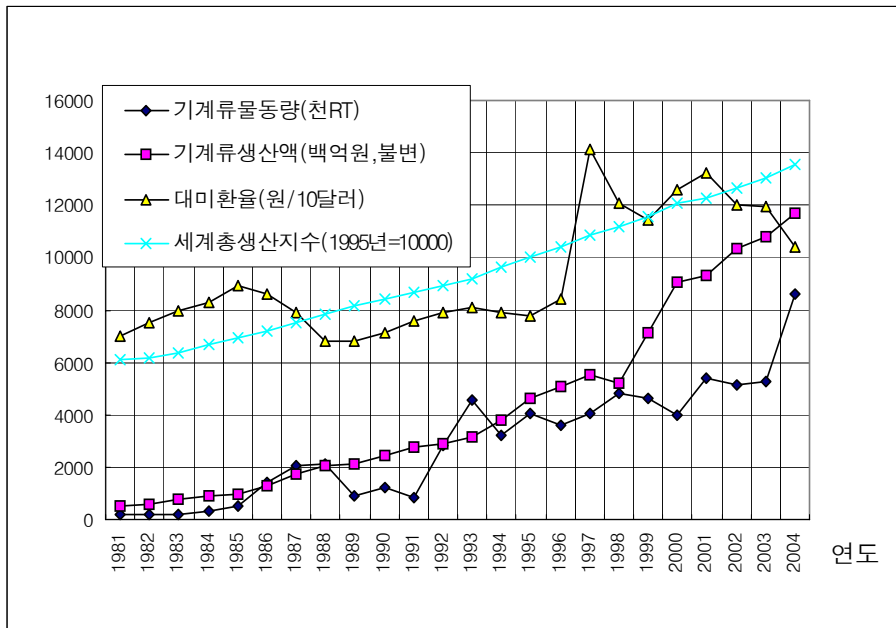
우리나라 일반 기계산업의 기술수준은 가공·조립기술에서 선진국에 상당히 근접한 것으로 나타나고 있으나 설계기술 면에서 상당한 격차를 보이고 있다. 구성 품목별로는 건설·광산기계, 농기계장비 등의 경우 선진국에 근접하거나 가까운 제품기술을 보유하고 있으나 여전히 핵심부품에 대한 해외의존도가 높다. 공작기계의 경우 가공·조립 수준의 일반 기술은 선진국에 접근해 있으나 설계기술, 제어기술 및 핵심부품개발 능력이 상당히 떨어지고 있다.

따라서 수출물동량은 대부분 가공·조립기술 위주의 기계류 품목이고 핵심부품이나 기술력은 해외에서 수입하고 있는 행태라서 기계류에 대한 수출을 늘릴수록 수입도 함께 늘어나는 교역구조를 이루고 있다. 이러한 이유로 물동량의 추세도 수입물동량과 비슷한 양상을 보이고 있으며 1993년에 단기 고점을 기록한 것과 이후 10년간 횡보추세를 나타낸 것도 거의 일치하고 있다. 동 제품의 물동량 추세는 같은 품목의 수입물동량처럼 약간은 불규칙하나 상승추세가 여전히 이어지는 ‘불규칙 상승형’ 흐름 종목으로 정의하였다.

〈그림 4-23〉 기계류 수출물동량 변동추이



〈그림 4-24〉 기계류 수출물동량 및 관련 변수 변동추이



(3) 차량 및 그 부품(상승 지속형)

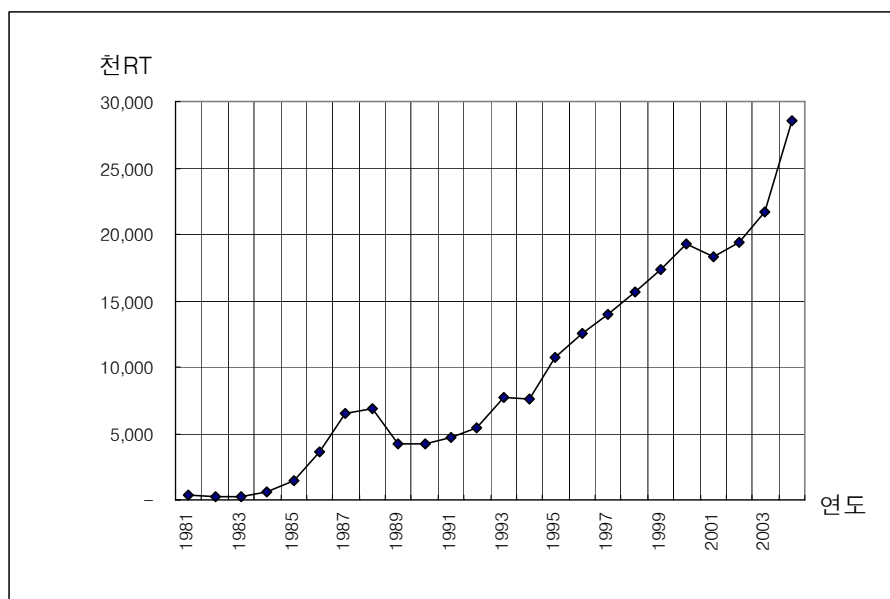
자동차 산업은 자본, 노동이 대규모로 투입되고, 다양한 기술분야가 결집된 산업으로서 관련산업에 대한 직접적인 전후방 효과가 매우 크다. 1960년대 초 KD 부품을 단순 조립하는 수준에서 출발하여 40여년 만에 개도국으로서는 유일하게 독자모형을 보유한 세계적인 자동차생산국으로 성장하였다.

자동차 수출물동량(R/T단위 항만물동량)은 1990년대 연평균 16.5%의 증가율을 기록하였고, 2000년대 들어서도 세계 자동차 시장의 전반적인 부진과 국내 내수 침체에도 불구하고 2004년까지 연평균 10.5%의 증가율로 상승추세를 이어오고 있다. 업체별로는 현대가 2003년 국내기업으로는 처음으로 100만 대 수출을 넘겼고, 기아는 서유럽 수출 급증을 통해, GM-대우는 2002년 10월 서유럽과 북미지역에 수출을 재개하게 됨으로써 수출물동량이 급증하였다. 해외 주요시장에서 국산차의 가격경쟁력 향상과 미국과 서유럽 내 판매망 확충,

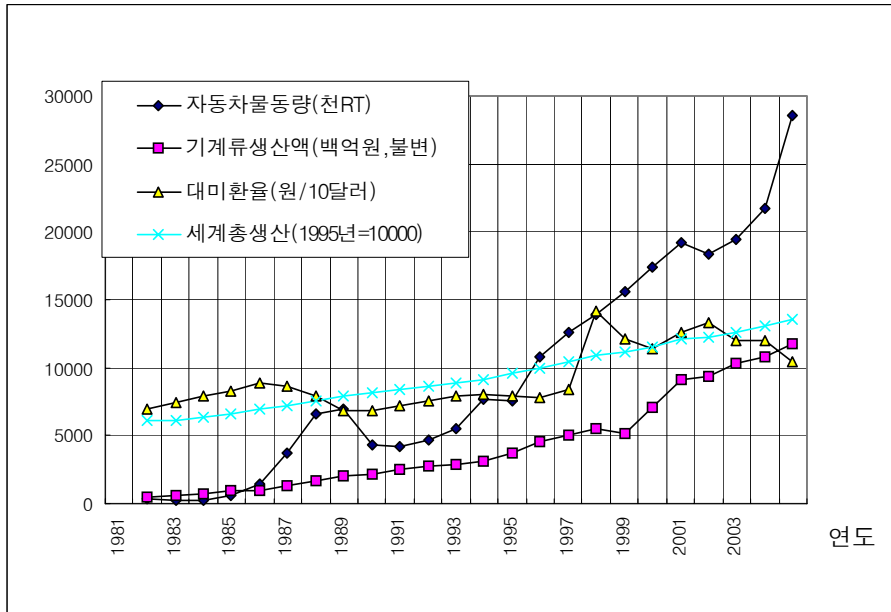
GM-대우의 수출 재개, 자동차 품질 개선 등에 힘입어 자동차 생산에서 수출이 차지하는 비중도 1980년 45%대에서 2003년 65%대를 넘어섰다.

자동차 수출물동량은 지속적인 상승흐름을 이어가고 있는 전형적인 ‘상승 지속형’의 변화추세를 보이고 있다. 그러나 2002년부터 감소 내지 정체 상태를 벗어나지 못하고 있는 내수 소비량, 생산기지의 해외 소비지로의 지속 이전 현상 등을 고려할 때 이러한 수출 상승세가 계속 이어질지는 미지수다. 자동차 수출물동량은 생산능력의 확충(기계류생산액으로 대변), 세계경기 동향, 가격 경쟁력을 표시할 수 있는 환율동향 등의 변수에 영향을 많이 받고 있다.

〈그림 4-25〉 자동차 수출물동량 변동추이



〈그림 4-26〉 자동차 수출물동량 및 관련 변수 변동추이

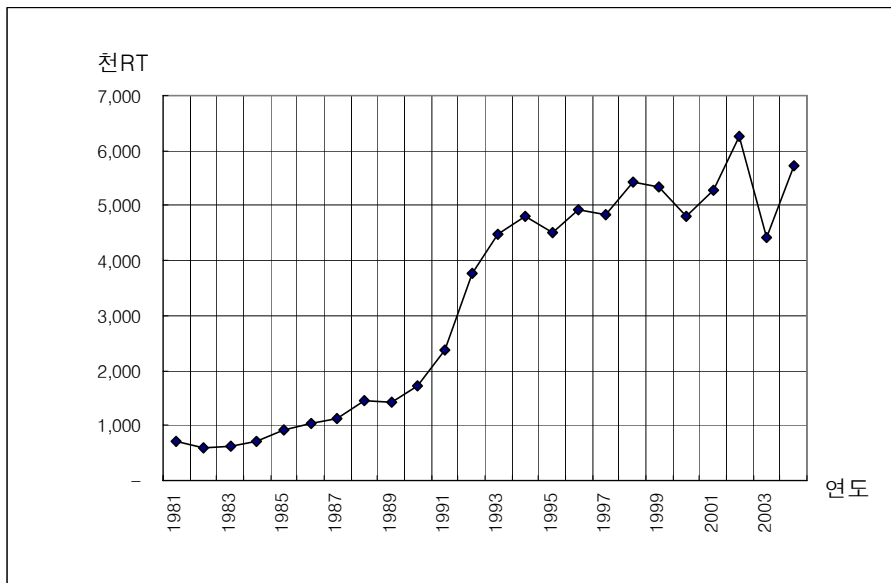


(4) 플라스틱/고무 및 그 부품(상승 둔화형)

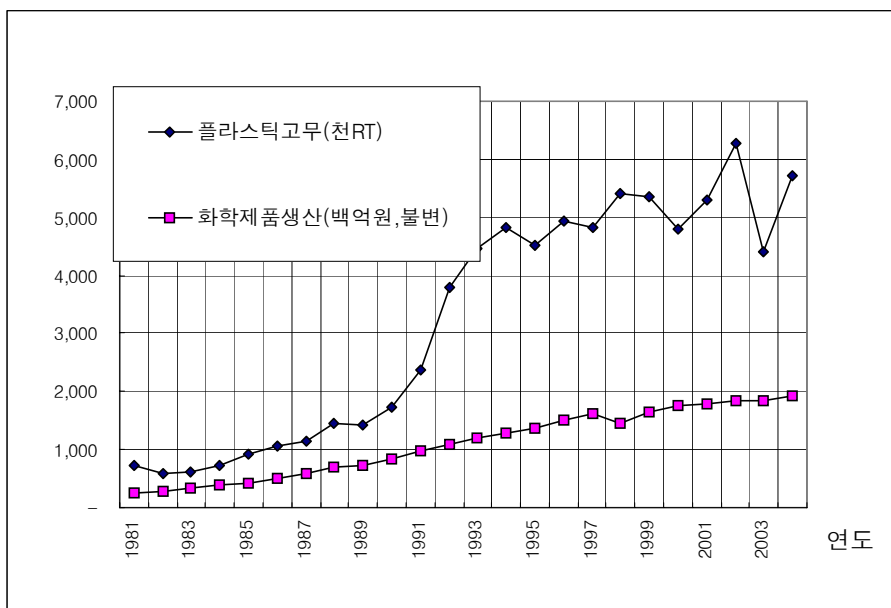
동 제품의 특성에 대해서는 수입물동량 부문에서 이미 상술한 바 있다. 동 제품의 수출물동량은 고무류 물동량의 60% 이상을 차지하고 있는 자동차 타이어의 수출 증가와 플라스틱 생산능력의 확충에 힘입어 '90년대 중반까지 비교적 큰 폭의 상승세를 이어왔다. 그러나 자동차, 전자, 건설 등 광범위한 산업분야에 투입되는 플라스틱류의 수출 강세에도 불구하고 신발, 고무, 타이어 등 전방 부문의 부진 지속으로 합성고무의 수출 상승세가 크게 꺾이며 전체 수출물동량의 증가세도 둔화되는 '상승 둔화형' 행태의 변화추이를 기록하고 있다.

동 제품은 관련 산업이라 할 수 있는 화학제품 생산액과 비교적 밀접한 상관관계를 이루고 있다.

〈그림 4-27〉 플라스틱고무 수출물동량 변동추이



〈그림 4-28〉 플라스틱고무 수출물동량 및 관련 변수 변동추이



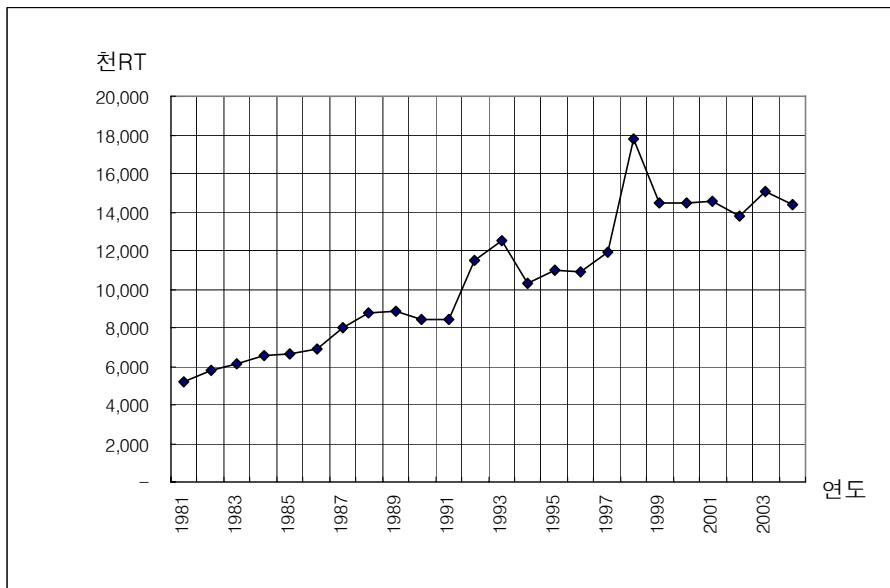
(5) 철강 및 그 제품(상승 둔화형)

동 품목은 수출이 국내 수주변동의 buffer로서 역할을 하여 철재의 국내수주가 감소할 때는 수출이 증가하는 반면, 국내수주가 증가하면 수출이 감소되는 특징을 보이고 있다. 즉, 수출되는 제품이 주로 토목, 건축용 자재로 사용되는 철근, 형강 등의 조강류 제품인데 국내 경기변동에 매우 민감하기 때문에 동 제품의 수출입에는 수급조절의 의미가 강하게 내포되어 있다고 할 수 있다.

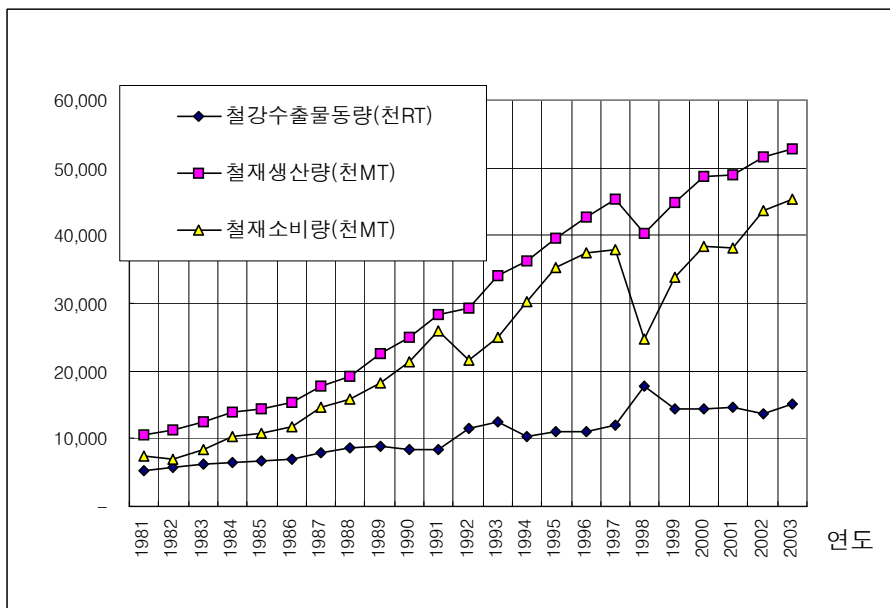
1970년대 동제품에 대한 수출은 급속히 늘어났으나 1980년대 들어서는 국내 경기팽창과 공업화 수요에 따른 철재의 국내 수요가 크게 증가함으로써 오히려 수출 증가세는 둔화되는 양상을 보였다. 1993년에 동 제품의 수출은 단기 고점을 형성한 뒤 이듬해부터 다시 회복세를 보이다 '98년에는 국내 경기의 침체로 인한 내수 악화와 환율 상승으로 인한 가격경쟁력 상승으로 수출물동량은 다시 사상 최고치를 기록하였다.

그러나 2000년대 들어 세계 철강경기 침체에 따른 국제 철강재 가격 하락과 세계적인 설비 과잉현상의 지속, 이로 인한 통상환경의 악화, 국내 설비능력 미증 등에 따라 수출물동량은 회복세를 지속하고 있다. 따라서 동 제품의 물동량은 시간이 갈수록 증가추세가 둔화되는 '상승 둔화형' 변화추이를 보이는 것으로 파악되었다.

〈그림 4-29〉 철강 수출물동량 변동추이



〈그림 4-30〉 철강 수출물동량 및 관련 변수 변동추이



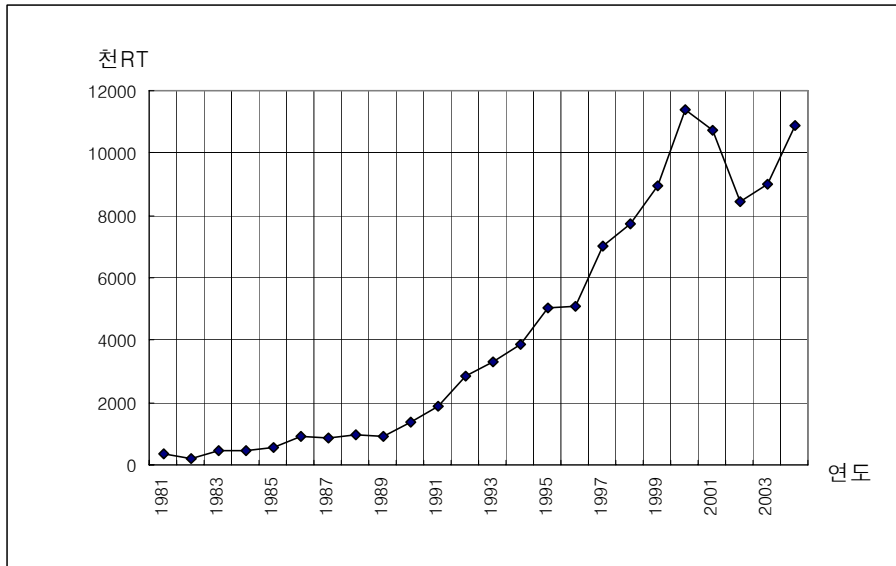
(6) 화학공업생산물(상승 둔화형)

동 제품은 수출 구성품목 중 중간재의 성격이 강한 유기화합물이 전체의 65% 이상을 차지할 정도로 편중되어 있어 수출 대상국의 공장 신설의 여건 변화에 큰 영향을 받고 있다. 수출대상지역은 2003년 기준으로 중국을 포함한 극동지역이 56%, 일본이 17%, 동남아 지역이 12%로 전체 물동량에서 아시아 지역의 비중이 85%를 넘고 있는데 특히 매년 큰 폭의 수출 증가율을 기록하고 있는 중국의 관련 산업군의 생산 활동에 큰 영향을 받고 있다.

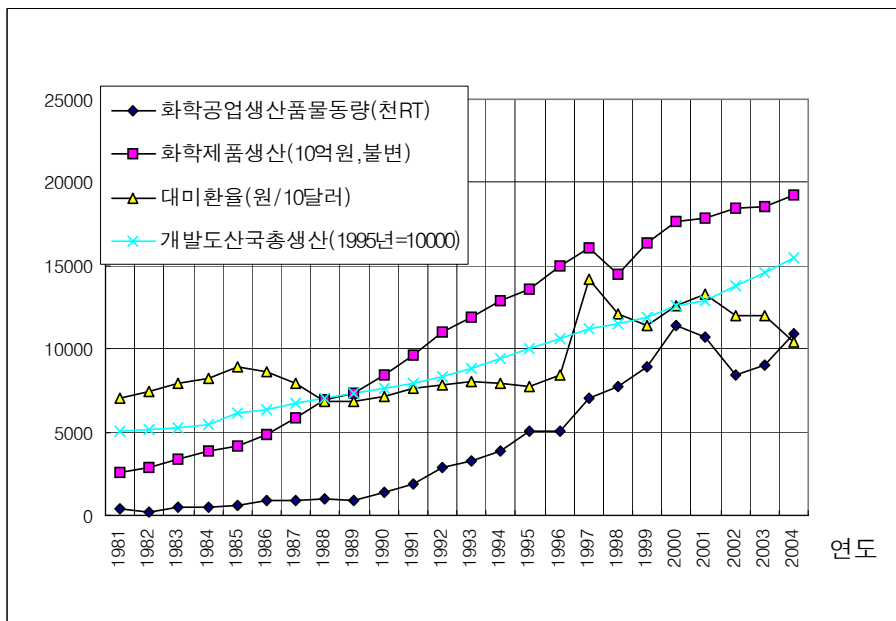
동 제품의 수출은 '80년대보다 중국과의 교역이 활발해지기 시작한 '90년대에 더욱 큰 폭의 물동량 증가율을 기록하고 있다. 2000년에 수출물동량이 고점에 이른 후 2001년과 2002년에 크게 감소한 것은 직전 연도에 비해 수출이 과도하게 폭증한 이유도 있겠으나 중국으로의 수출이 둔화된 것이 직접적인 원인으로 지적되고 있다. 그러나 중국의 경제 성장세가 지속되고 신흥시장의 하나인 동남아 지역의 경기 활황과 일본의 경기회복세에 힘입어 수출물동량은 다시 큰 폭의 회복세를 보이고 있다.

'90년대 이후 급속한 증가율을 기록한 동 제품의 수출물동량은 차츰 그 증가세가 둔화되는 '상승 둔화형' 패턴을 보이고 있는 것으로 파악되었다.

〈그림 4-31〉 화학공업생산물 수출물동량 변동추이



〈그림 4-32〉 화학공업 수출물동량 및 관련 변수 변동추이

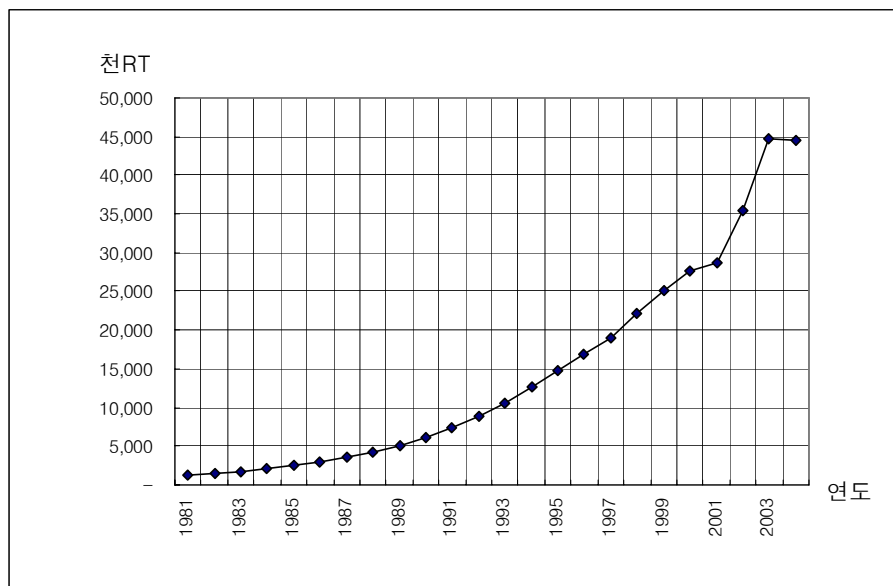


(7) 방직용섬유 및 그 제품(상승 지속형)

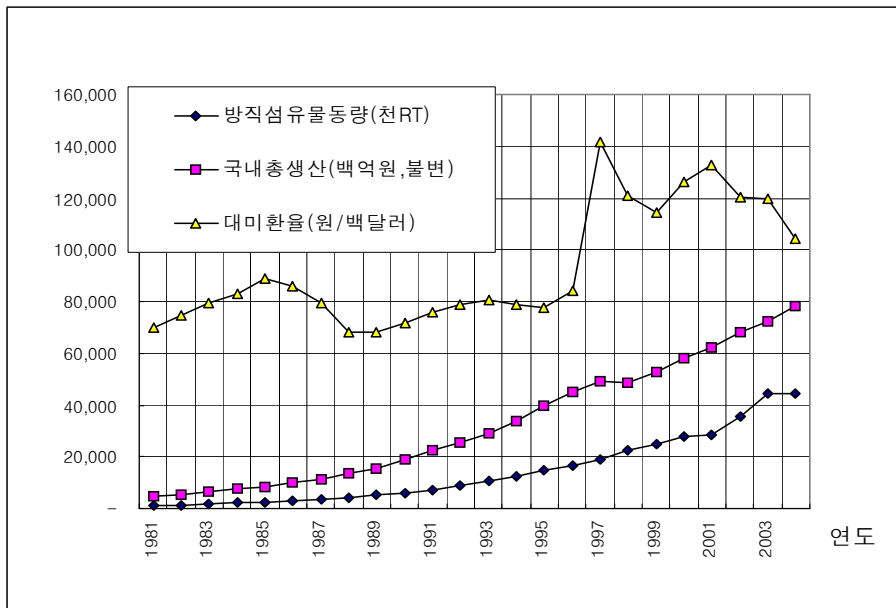
우리나라는 섬유산업을 통해 매년 100억 달러 이상의 무역흑자를 기록하였다. 섬유산업은 최근 각국의 수입규제, 중국 등 동남아 국가와의 경쟁심화 등 수출 장애요인이 산재해 있으나 합성방직과 직조기술이 뛰어나고, 특히 기능성 및 산업용사와 직물은 경쟁력을 계속 유지하고 있는 상태이다. 미국은 한국이 전 세계에 원단을 공급하는 주요 국가로서의 역할을 지속할 것으로 내다보고 있으며 중국으로 생산기지를 이동시키면서도 남자용 드레스셔츠, 드레스, 그리고 패션의류 등 특정 품목에 대해서는 계속 외주를 주고 있다.

견실한 생산기반과 기초원료를 제외한 원부자재의 자급능력, 세계 2위의 쿼터 보유량 등 지속적인 수출 증가요인을 바탕으로 '80년대 이후 꾸준한 상승세를 이어오고 있다. 2000년대 들어서도 연평균 두 자릿수의 증가율을 유지하고 있는 '상승 지속형' 물동량 추세를 보이고 있다.

〈그림 4-33〉 방직용섬유 수출물동량 변동추이



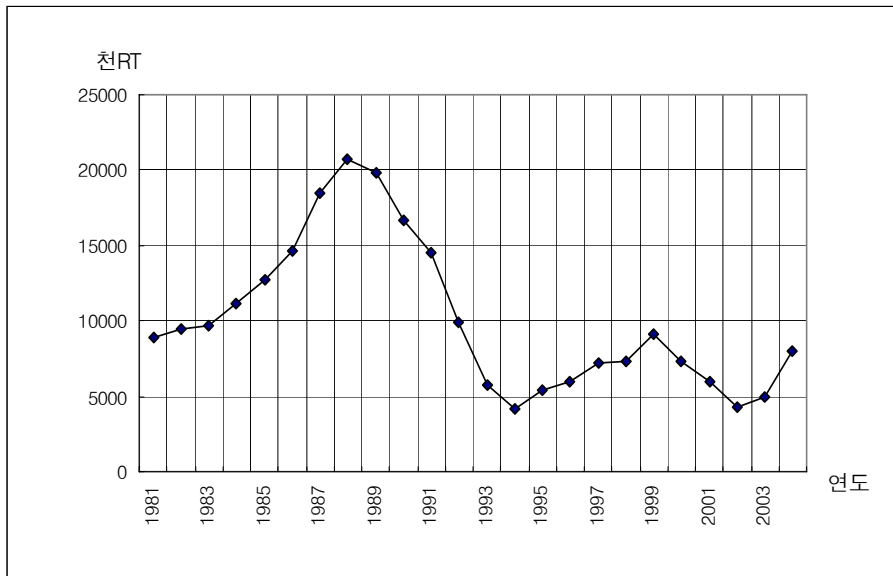
〈그림 4-34〉 방직용섬유 수출물동량 및 관련 변수 변동추이



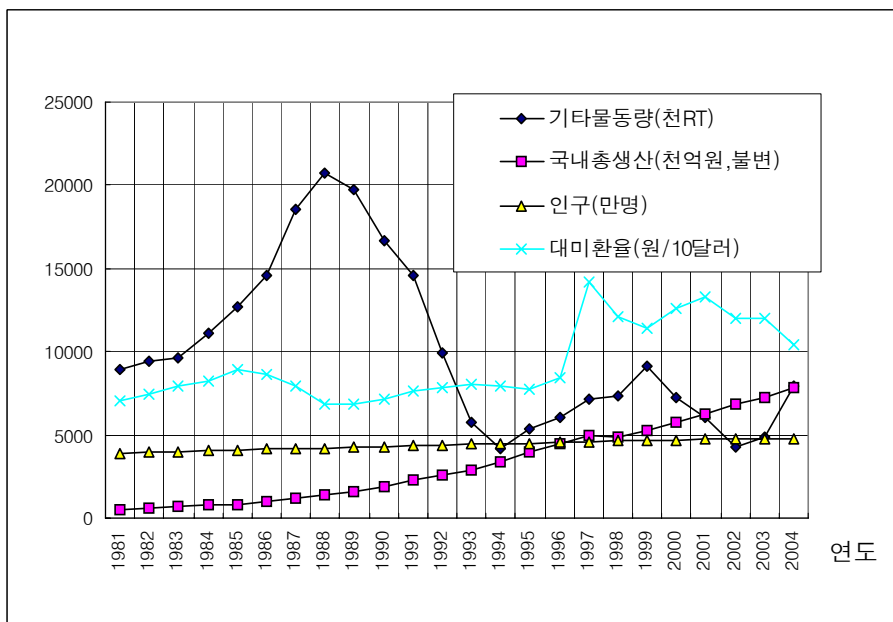
(8) 기타(정체형)

동 제품은 펄프, 판지, 우산, 도자기, 유리, 완구, 석면 등 생산에 필요한 기술수준이 낮은 품목들로 구성되어 있어 인건비나 산업구조적인 측면에서 수출물동량의 정체 혹은 감소세는 피할 수 없다. 1988년을 최고점으로 '90년대 중반까지 물동량이 급락하였으나 이후 500만 톤을 중심으로 등락을 반복하는 '정체형'의 물동량 변화추이를 보이고 있다.

〈그림 4-35〉 기타 수출물동량 변동추이



〈그림 4-36〉 기타 수출물동량 및 관련 변수 변동추이

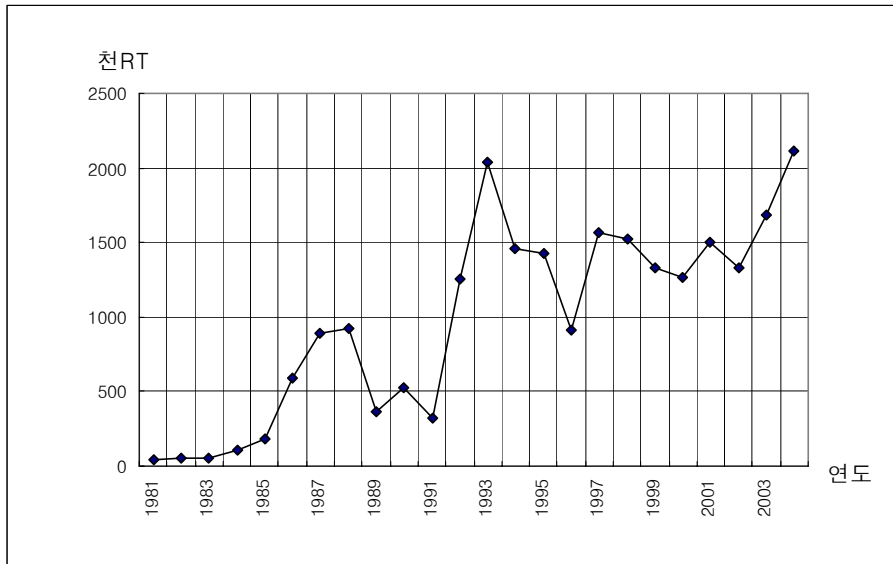


(9) 비철금속 및 그 제품(상승 둔화형)

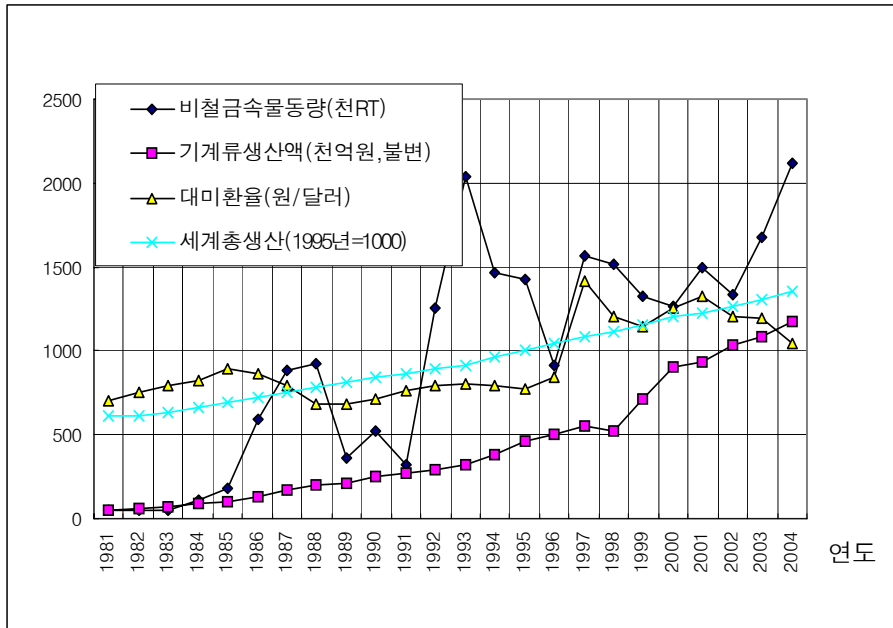
우리나라는 주요 비철금속 품목의 생산량과 소비량에서 세계 10위권을 유지하고 있다. 최근 후방산업인 자동차, 전자, 중화학공업의 경기가 호조를 보이고 있고 생산능력의 대폭 확충은 물론, LME 지정창고를 광양에 유치함으로써 '90년대 중반 이후 정체 상태를 보이던 동 제품에 대한 수출물동량은 2003년과 2004년에 연속하여 증가세를 나타내고 있다.

'93년 고점을 기록한 이후 일정한 범위 내에서 상승과 하락 추세를 반복하다 2004년에 '93년의 고점을 갱신하였다. 그러나 비철금속 산업이 공해 유발산업으로 분류되어 있어 세계적으로 환경규제가 심하게 적용되고 있고, 관련 산업의 외국 현지 생산이 증가하고 있어 수출물동량이 지속 증가추세를 보이기 힘든 '상승 둔화형' 형태를 보이는 것으로 판단되었다.

〈그림 4-37〉 비철금속 수출물동량 변동추이



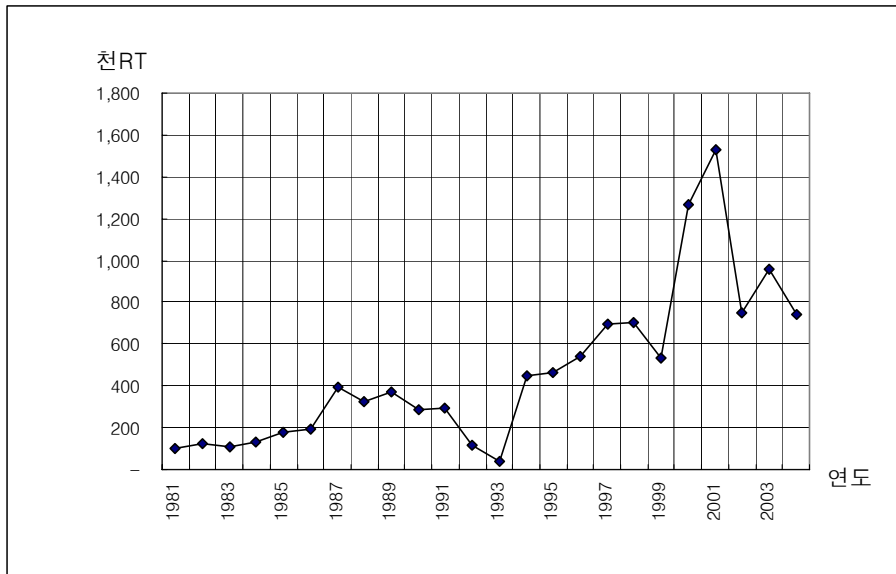
〈그림 4-38〉 비철금속 수출물동량 및 관련 변수 변동추이



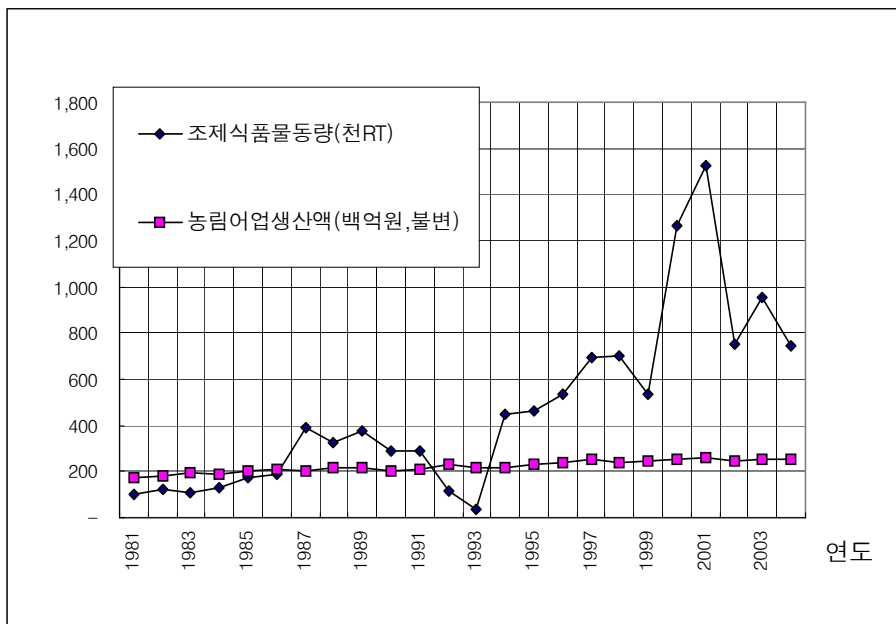
(10) 조제식품, 음료, 주류 등(상승 둔화형)

동 제품의 주요 품목은 수산물 가공품, 면류, 각종 조제식품 등으로 수입에 비해 물동량이 상대적으로 적은 편이다. 1990년대 초와 1999년에 일시적인 조정을 보이긴 했지만 2001년 최대 수출량을 기록할 때까지 꾸준한 상승세를 이어왔다. 그러나 2002년 큰 폭의 물동량 감소에 이어 이후 거의 정체 상태를 나타내고 있다. 주요 품목 중 수산물 가공품의 경우 최대 수요처인 일본의 소비수요 감소와 위생관리의 강화, 중국 등 동남아 산의 저가공세 및 주요 생산어종의 생산 감소로 수출 정체 현상을 보이고 있다. 조제식품은 기술의 발전과 주변국 수요 증대로 수출이 늘어날 것으로 보이나 수산물의 확보가 어렵고 원료 대부분을 수입에 의존해야 하기 때문에 역시 증가에 한계가 있을 수밖에 없다. 동 제품의 수출물동량은 전반적인 상승추세를 이어나가긴 하였으나 업종구조상 증가추세가 둔화될 수밖에 없는 ‘상승 둔화형’ 추이를 보이고 있다.

〈그림 4-39〉 조제식품 수출물동량 변동추이



〈그림 4-40〉 조제식품 수출물동량 및 관련 변수 변동추이



제5장

품목별 예측의 적합도 비교·분석

1. 방법론

예측의 적합도란 모형에 의한 예측 결과가 당해연도의 실적치에 얼마나 잘 부합되는가를 의미한다. 이번 소절에서는 적합도의 비교·분석을 기본적으로 회귀분석과 신경망분석을 대상으로 실시하되, 추가적으로 본 원이 기존 연구보고서²⁹⁾에서 사용했던 변형된 형태의 회귀예측모형과 신경망분석의 비교 연구도 함께 수행하였다. 변형된 회귀모형이란 종속변수에 해당하는 품목별 물동량의 변화에 대한 독립변수의 설명력을 높이기 위해 종속변수 혹은 설명변수를 로그함수 혹은 지수함수 형태로 변수 변환하거나, 선형함수의 형태를 비선형 형태로 치환한 것을 뜻한다.

모형별 예측의 적합도 비교는 기본적으로 품목별로 이루어지나 비교·분석에 근거한 데이터 마이닝 기법의 유용성 평가를 위해 제3장에서 구분된 4가지 물동량의 변화유형(성장지속형, 불규칙 성장형, 성장 둔화형, 정체 혹은 불규칙 변화형)을 분석의 기본으로 삼았다. 적합도 비교는 각 품목별로 1981~2000년까지의 관찰개수를 사용하여 각각의 모형을 추정한 뒤 2001~2004년의 4개 년도에 대해 추정모형에 의한 예측치와 실적치를 대비시켜 모형별 적합도를 비교·판정하였다. 이 때 비교·판정의 기준을 위한 통계적 특성치는 MAE(Mean Absolute Error)로 설정하였고 R^2 (결정계수)를 참고 지표로 이용하였다. 각 품목별 물동량 예측에 사용된 설명변수는 별도의 선정과정 없이 기 발간된 물동

29) 해양수산부, 「전국 항만물동량 예측」, 2004. 12.

량 예측보고서³⁰⁾의 것을 그대로 인용하였으며 그 관계는 <표 5-1>에 잘 나타나 있다.

$$MAE = \frac{\sum_{i=1}^N |\hat{y}_i - y_i|}{N}$$

$$R^2 = \frac{N \left[\sum_{i=1}^N \hat{y}_i y_i \right] - \left[\sum_{i=1}^N \hat{y}_i \right] \left[\sum_{i=1}^N y_i \right]}{\sqrt{\left[N \sum_{i=1}^N \hat{y}_i^2 - \left(\sum_{i=1}^N \hat{y}_i \right)^2 \right] \left[N \sum_{i=1}^N y_i^2 - \left(\sum_{i=1}^N y_i \right)^2 \right]}}$$

N : 관찰 개수

y_i : 품목별 물동량

\hat{y}_i : 품목별 예측 물동량

30) 해양수산부, 「전국 항만물동량 예측」, 2004. 12

2. 적합도 비교·분석

본 소절에서는 물동량의 4가지 추세유형별 혹은 품목별로 각기 예측모형의 적합도에 있어 어떠한 차이점을 보이는지 살펴보았다. 즉, 각 유형별·품목별로 MAE와 R^2 를 비교해 보고 각기 우열한 모형은 무엇인지, 추세 유형별로 공통점은 존재하는지 그리고 그 원인은 무엇인지를 통계적 특성치와 <그림>을 통해 분석하였다. 물론 추세유형과 상관없이 예외적인 경우에 대해서도 그 원인을 파악하려 노력하였다.

1) 회귀모형과 신경망모형 분석의 결과 비교

회귀모형과 신경망모형 간의 적합도 비교결과 추세유형별로 확연히 구분되는 특징을 보여주고 있다. 수입 유연탄을 비롯해 ‘상승 지속형’에 속하는 7개 품목 중 수입 방직섬유를 제외한 6개 품목의 예측 적합도에서 회귀모형이 신경망모형보다 우수한 것으로 나타났다. 수입 유연탄의 경우 회귀모형에 의한 예측치의 MAE는 신경망모형의 25%에 불과하였고 R^2 는 0.996으로 신경망모형의 0.927을 훨씬 상회하고 있다. 특히 ‘상승 지속형’의 대표적인 추세변동 흐름을 보이고 있는 수입 기계류는 MAE가 신경망모형의 약 1/5 수준이며 R^2 는 0.987 대비 0.767로서 회귀모형의 설명력이 신경망모형에 비해 월등한 모습을 보여주고 있다. ‘상승 지속형’에 속하는 7개 품목을 모두 합산할 경우에도 신경망모형의 MAE는 회귀모형의 1.28배에 해당함으로써 적합도 면에서 신경망모형에 비해 우수하게 나타났다.

수입 비철금속을 비롯한 ‘불규칙 상승형’의 품목은 3개 중 2개 품목에서 회귀모형이 상대적으로 양호한 것으로 평가되고 있다. 수출 기계류와 수출 전기기기의 경우 회귀모형의 MAE가 신경망모형에 비해 각각 1/2과 1/3에 불과하고 R^2 도 각각 0.963 및 0.985로서 신경망모형에 비해 10%포인트 이상 설명력

이 더 높은 것으로 분석되었다. 이와는 반대로 수입 비철금속의 예측모형은 신경망모형이 더 나은 것으로 나타났으나 ‘불규칙 상승형’의 품목 전체를 함께 고려할 경우 회귀모형의 MAE는 신경망모형 대비 1:1.44로 나타나 역시 동 유형의 전반적인 예측모형은 회귀모형이 적합도 면에서 신경망모형보다 더 나은 것으로 파악되고 있다.

‘상승 둔화형’의 경우 앞의 두 유형과는 정 반대의 결과를 낳고 있다. 수입 전기기기를 비롯 7개 품목이 이 유형에 해당되는데 수입 전기기기와 수출 철강제품을 제외한 5개 품목에서 신경망모형이 회귀모형보다 더 적합도가 높은 것으로 나타났다. 대표적인 예로 수입 화학공업제품은 신경망모형의 MAE가 회귀모형의 22% 수준밖에 되지 않으며, 회귀모형에서 0.677에 불과하던 R^2 는 신경망모형에서의 0.978로 대폭 호전되고 있다. 우리나라 컨테이너물동량의 상당부분을 차지하고 있는 수출 화학공업과 수출 플라스틱 제품의 경우도 통계적 특성치들이 신경망모형에서 대폭 호전되어 나타나고 있다. ‘상승 둔화형’에 해당되는 모든 품목의 MAE를 단순 합산할 경우 신경망모형의 MAE는 회귀모형의 66%에 불과함으로써 동 유형에 속한 품목별 물동량의 예측시 회귀모형보다는 신경망모형을 사용하는 것이 예측의 적합도를 높일 방편이 될 수 있다.

수입 플라스틱·고무제품, 기타 수출입 잡화 품목의 물동량은 추세를 가름하기 힘든 흐름을 보이고 있다. 본 서에서는 이러한 ‘정체형 혹은 불규칙형’의 추세 유형에 해당되는 품목을 전술한 3가지로 분류하고 있으나 산업의 규모가 커지고 그 구조도 급격히 변화되고 있는 시대 흐름에 비추어 이와 같은 추세를 보이는 품목의 수는 차츰 늘어날 것으로 전망된다. 따라서 이러한 품목에 대한 물동량을 어떠한 함수모형을 통해 계량적인 접근을 통해 예측한다는 것은 불가능하거나 아주 어려운 작업이 될 수 있다. 본 서에서는 동 유형에 속한 품목이 3개밖에 되지 않아 품목별로 적합도에 대한 객관성 있는 판정을 내리는 것이 별로 유의하지 않다고 판단하고 있다. 다만 전체적으로 볼 때 선형적인 함수모형을 기본 바탕으로 하는 회귀모형보다는 불규칙한 물동량의 흐름 속에서 비선형적인 관계식을 통해 나름대로의 정형화된 특성을 찾아나가는 신

경망모형이 더 적합할 수 있다는 판단이다. 물론 동 유형에 속한 3개 품목의 통계적 특성치들을 모아 놓고 보아도 신경망모형이 회귀모형보다 우수한 것으로 나타났다.

〈표 5-2〉 예측의 적합도 비교분석 I (Regression과 Neural Network)

추세 유형	품목	MAE		R-square		MAE의 상대비율 ¹⁾	적합모델
		Regression	Neural Network	Regression	Neural Network		
상승 지속형	수입_유연탄	35.75	143.73	0.996	0.927	4.02	Regression
	수입_기타광석	20.52	24.80	0.976	0.959	1.21	Regression
	수입_기계류	5.79	24.12	0.987	0.767	4.17	Regression
	수입_철강	39.22	42.77	0.936	0.923	1.09	Regression
	수입_방직섬유	116.52	67.08	0.728	0.875	0.58	Neural Network
	수출_차량	34.62	34.98	0.975	0.946	1.01	Regression
	수출_방직섬유	76.40	83.26	0.948	0.935	1.09	Regression
	소계	328.83	420.74			1.28	
불규칙 상승	수입_비철금속	13.35	5.15	0.667	0.937	0.39	Neural Network
	수출_기계류	8.55	17.74	0.963	0.866	2.07	Regression
	수출_전기기기	6.56	18.05	0.985	0.889	2.75	Regression
	소계	28.46	40.93			1.44	
상승 둔화	수입_전기기기	4.69	5.48	0.908	0.845	1.17	Regression
	수입_화학공업	54.65	12.54	0.677	0.978	0.23	Neural Network
	수출_조제식품	3.48	3.43	0.788	0.798	0.99	Neural Network
	수출_비철금속	4.04	3.29	0.929	0.936	0.81	Neural Network
	수출_철강	6.28	24.06	0.997	0.969	3.83	Regression
	수출_화학공업	29.16	19.09	0.896	0.951	0.65	Neural Network
	수출_플라스틱고무	10.09	6.40	0.951	0.982	0.63	Neural Network
	소계	112.38	74.27			0.66	
정체	수입_플라스틱고무	3.96	2.98	0.834	0.903	0.75	Neural Network
	수입_기타	2.79	9.65	0.996	0.958	3.46	Regression
	수출_기타	98.18	13.52		0.936	0.14	Neural Network
	소계	104.93	26.15			0.25	

주 : 1) MAE of NN / MAE of Reg.

이와 같이 지속 상승이던 불규칙 상승이던 물동량의 상승이 계속되고 있는 것으로 평가된 품목의 경우 선형관계를 바탕으로 한 회귀모형이 전반적으로 적합도 면에서 우수한 것으로 나타났다. 반면 상승이 둔화되거나 정체 혹은 추

세를 판단할 수 없을 정도의 물동량 변화를 보이고 있는 품목은 대부분 과거 급속한 경제발전을 주도해 왔던 제조업 위주의 산업구조와 직·간접적으로 밀접한 관련이 있을 가능성이 높은 것들이다. 이러한 품목들은 일정한 추세를 전제로 한 회귀분석보다는 모형에서 필요한 분포의 가정을 통하여 가상의 생성 데이터를 통해 새로운 예측 상황을 만들어 가는 신경망모형이 보다 타당한 결과를 낳을 수 있다.

그러나 이러한 추세유형별 일반론에도 불구하고 전혀 다른 결과를 보이는 품목들이 있다. ‘상승 지속형’에서의 수입 방직섬유와 ‘불규칙 상승형’의 수입 비철금속, 그리고 ‘상승 둔화형’의 수입 전기기기 및 수출 철강이 그 품목들이다. 본 서에서는 신경망모형의 특성을 토대로 나름대로 그 이유를 분석해 보려 한다. 물론 충분한 근거 사유를 찾기가 어려울 뿐더러 신경망모형에서는 초기 예측 과정에서부터 최종 예측 결과에 이르기까지 사용된 함수를 전혀 알 수 없기 때문에 예측 결과에 대한 해석이 불가능하다는 점이 더욱 그러하다. 따라서 본 서에 기술된 내용은 일반론에 근거한 추론적 수준일 수 있음을 상기할 필요가 있다.

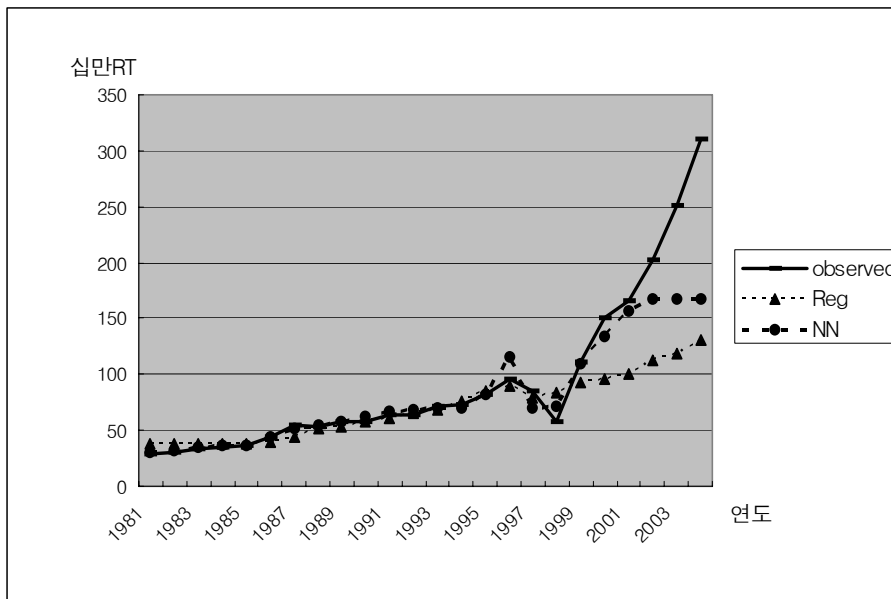
수입방직섬유는 예측에 사용된 자료 초기 시절보다 시간이 갈수록 더욱 가파른 증가세를 보이고 있다. 이러한 점은 선형 형태의 ‘상승 지속형’을 보이는 여타 품목들과는 약간 다르다. 즉 ‘상승 지속형’이 지속 상승을 하지만 선형 형태를 보일 경우 실질적인 물동량의 증가율은 물동량의 규모가 커질수록 조금씩 낮아진다는 점이다. 특히 수입방직섬유의 경우 우리나라가 겪었던 '90년대 후반 환란위기를 넘기면서 더욱 높은 증가세를 보이고 있는데 회귀모형이 '80년대 이후 과거의 실적을 골고루 반영하여 완만한 흐름을 보이고 있는 반면 신경망모형에서는 '90년대 후반 이후의 높은 물동량 증가세에 크게 영향을 받고 있음을 지적할 수 있다.

‘불규칙 상승형’에서의 수입 비철금속의 경우 수입방직섬유와는 상당히 대조되는 물동량 변화추이를 보이고 있다. 수입 비철금속은 '80년대 이후 지속 상승추이를 보이다 '94년에 고점을 기록한 이후 거의 10년간 횡보세(상승과 하락

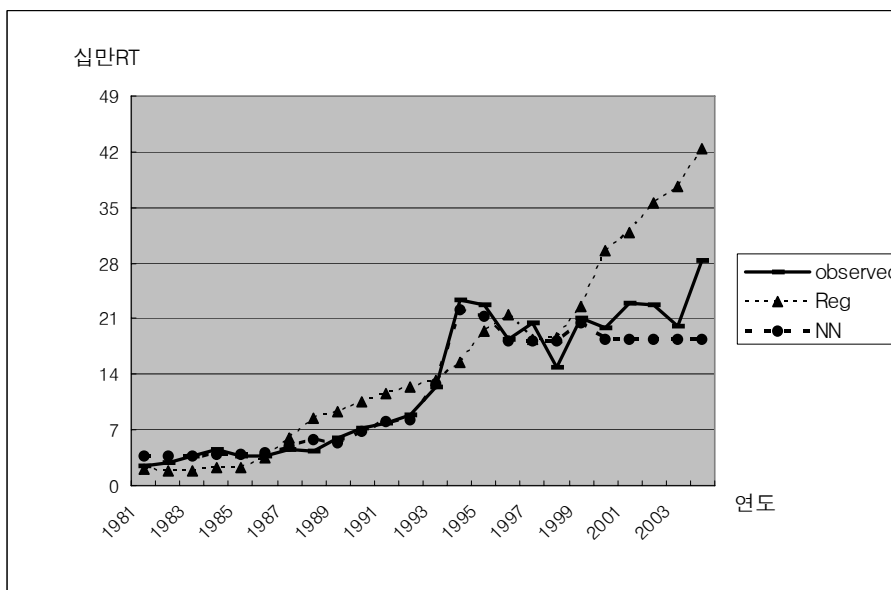
반복)를 계속하였다. 이후 경기회복과 LME 기지창고 지정을 계기로 횡보세에서 탈피하여 10년 만인 2004년에 다시 최고치를 기록하였다. 수입 비철금속은 2004년의 급격한 실적 상승이 없었다면 ‘상승 둔화형’ 형태의 물동량 움직임을 보인다고 해도 틀림없을 것이다. 회귀모형에 의한 예측 결과가 제대로 실적치에 부합되지 않는 이유도 여기서 찾을 수 있다. 회귀모형에 의한 추정치는 '80년대 중반 이후 고점을 기록한 '94년까지의 급격한 물동량 상승률에 크게 영향을 받고 있는 반면 신경망모형은 예측하기 어려운 2004년을 제외하고는 '90년대 중반 이후의 물동량 변동을 정확히 반영하고 있다.

‘상승 둔화형’의 예외 사례인 수입 전기기기는 수입 비철금속의 경우와 대비시켜 살펴보면 이해하는 데에 도움이 될 수 있다. 두 품목 모두 상승이 둔화되는 형태를 보이고 있으나 수입 전기기기의 물동량 조정폭이 상대적으로 깊고 수입 비철금속과는 달리 회복세가 강하면서도 빠르지 못하다는 점이 동 유형에 포함된 사유라 할 수 있다. 그러나 조정기간이 환란에 따른 금융위기 이후 2000년대 초반까지로 짧다는 점과 전 고점을 역시 돌파했다는 점에서 불규칙적이거나 상승형에 가까운 추세 유형에 속한다는 점이 인정될 수 있다. 물론 본서에 품목별 물동량 추세유형의 구분 기준이 엄밀하게 수립되어 있지 않다는 점이 이러한 예외 사례를 초래한다는 점도 인정하지 않을 수 없다. 그러나 예외적인 경우에도 해당 품목의 물동량 추세유형에서 많은 공통적인 사유를 발견할 수 있었고, 회귀모형과 신경망모형의 적용 가능한 범위에 대해 시사 가능한 여러 요인들이 파악되었다.

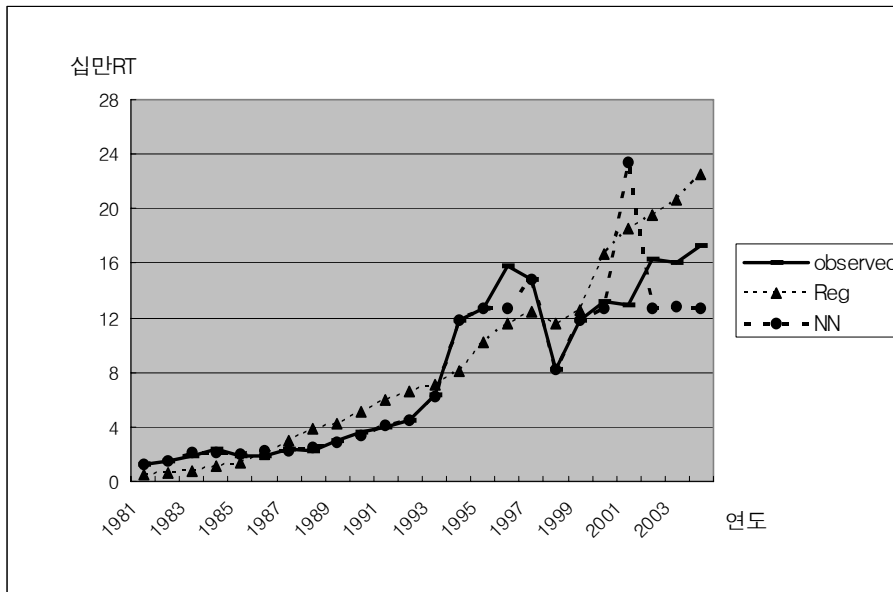
〈그림 5-1〉 방직용섬유 및 그 제품(수입)의 실적치 및 예측치 비교



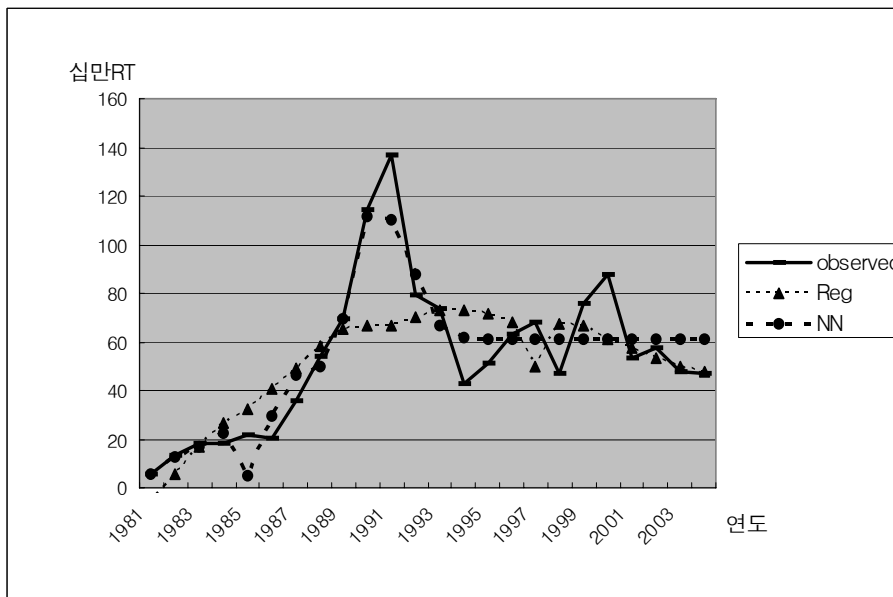
〈그림 5-2〉 비철금속 및 그 제품(수입)의 실적치 및 예측치 비교



〈그림 5-3〉 전기기기 및 그 제품(수입)의 실적치 및 예측치 비교



〈그림 5-4〉 기타(수입)의 실적치 및 예측치 비교



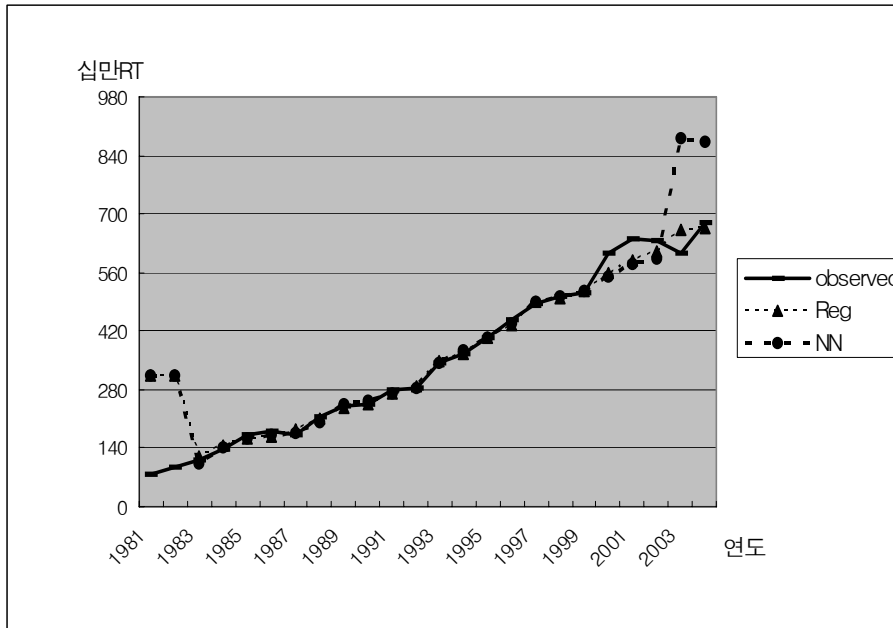
그 같은 점들은 물동량의 추세유형별로 일반론적 추론이 잘 들어맞는 품목에서도 찾아 볼 수 있다. 다음의 <그림>들은 추세유형별로 적합도가 높은 예측모형이 정해져 있다는 추론을 했을 때 이에 가장 잘 부합되는 몇 가지 품목의 예측치와 실적치를 대비시킨 것이다. 수입 유연탄은 선형 형태의 지속적인 상승세를 보여 왔다. 물동량의 증가세가 과거에 비해 큰 폭의 변화를 보였던 2000년의 경우를 제외하곤 예측모형에 상관없이 예측치는 비교적 실적치에 잘 부합되었다. 그러나 2000년의 물동량 폭증세는 과거 실적자료 전체를 고루 반영하는 특성을 가진 회귀모형보다는 최근의 물동량 움직임에 비교적 민감하게 반응하는 신경망모형에 더 큰 영향을 미쳤다. 수입유연탄은 2004년 다시 과거의 추세를 회복하였고 회귀모형은 결과적으로 실적치에 더 잘 부합되게 되었다.

‘불규칙 상승형’의 수출 전기기기 제품은 상승 및 하락폭이 너무 크다는 특징이 있으나 일정 기간의 상하 변동기간을 거친 뒤 저점을 한 단계씩 높여가는 전형적 불규칙 상승형의 움직임을 보여주고 있다. '92년과 '93년 연속으로 큰 폭의 상승을 기록하였으나 이후 관찰치의 마지막 기간인 2000년까지 500만~600만 톤 사이를 움직이며 정체 상태를 벗어나지 못했다. 따라서 신경망모형의 경우 최근의 실적치의 변동을 크게 반영하여 '90년대 중반 이후 물동량의 증가세가 정체 상태에 가깝게 둔화되는 양상을 예측하였고 회귀모형은 '80년대에서 '90년대 초반까지의 증가세 영향을 고루 반영함으로써 완만한 선형형태의 예측을 하였다. 그러나 실제 물동량은 2004년 다시 급속한 수출 신장세에 힘입어 1993년의 고점을 갱신하였고 1981년과 2004년의 물동량 실적을 연결하면 결국 장기적으로는 회귀모형에서 예측했던 것처럼 선형의 완만한 상승세를 기록함으로써 회귀모형의 적합도를 높여주게 되었다.

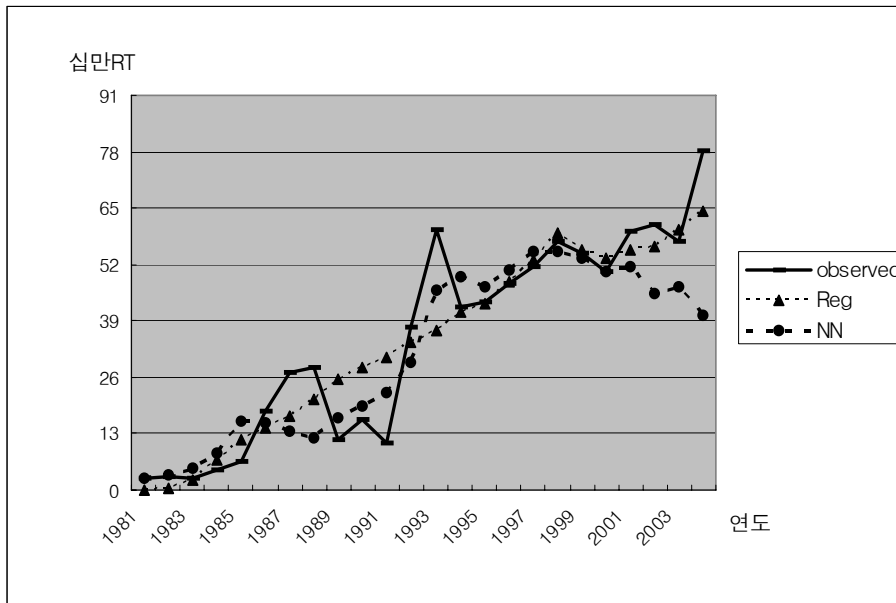
‘상승 둔화형’의 수출 플라스틱·고무 품목은 '90년대 중반까지 급격한 증가율을 보여 왔으나 이후 상승이 급격히 둔화되었고 2000년까지 500만 톤을 중심으로 폭이 아주 작은 상하 등락을 반복하였다. 이러한 단기적인 물동량의 흐름은 신경망모형에서 더욱 잘 반영되었고 그에 따른 예측치도 거의 정체 상태에

가까운 완만한 상승세로 나타났다. 물론 2001년 이후 실제 물동량의 변화도 과거와 양상이 약간 다르긴 하였으나 상하 등락을 반복하였고 따라서 지속 상승세를 예측한 회귀모형보다 신경망모형의 예측 결과에 훨씬 근접한 수치를 기록하였다.

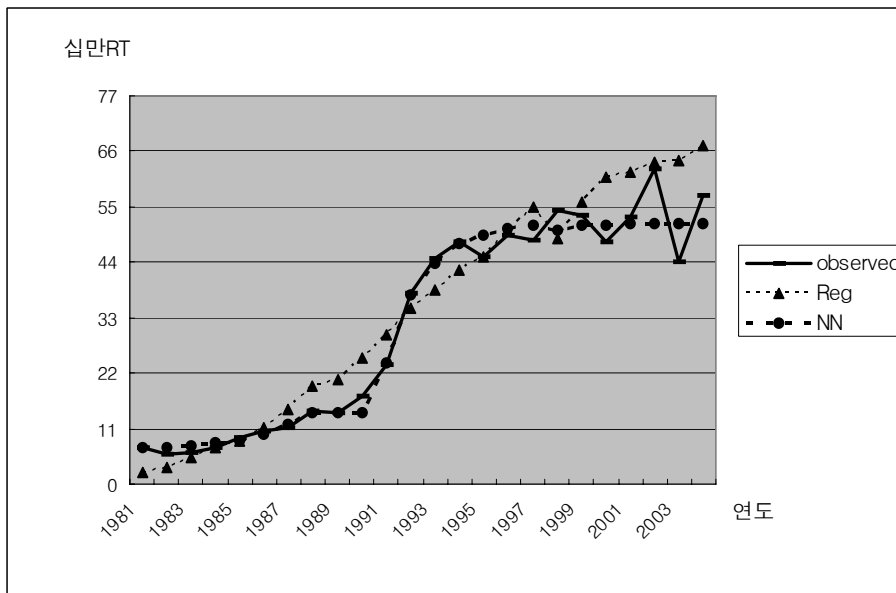
〈그림 5-5〉 유연탄(수입)의 실적치 및 예측치



〈그림 5-6〉 전기기기 및 그 부품(수출)의 실적치 및 예측치



〈그림 5-7〉 플라스틱고무 및 그 부품(수출)의 실적치 및 예측치



2) 변형회귀모형과 신경망모형 분석의 결과 비교

본 서에서는 단순한 선형 형태의 회귀모형과 신경망모형을 비교하였을 뿐 아니라 기존에 사용 가능한 여러 변형된 형태의 회귀모형과 신경망모형과의 비교·분석을 수행하였다. 기존에 시도된 회귀모형의 변형 형태는 다양한 예를 들 수 있겠으나, 그 범위가 너무 넓어 본 서에서는 변수 선정을 위해 인용하였던 2004년 본 원의 용역연구보고서인 「전국 항만물동량 예측」상의 품목별 회귀예측모형을 그대로 적용하였다. 변형회귀모형에 대한 정의는 이미 기술한 바 있는데, 통상적으로 모형의 추정을 위해 사용될 관찰치의 최근 변동이 심각하거나, 정성적 선행 분석결과 향후의 예측치가 과거와는 구조적으로 달리 나타나야 하는데 이를 반영해야 할 경우 사용된다.

변형회귀모형과 신경망모형에 의한 예측 결과의 적합도 비교는 앞서 수행된 분석과정이나 절차가 같은 데다 그 결과에 있어서도 앞선 분석결과와 물동량의 추세유형별로 크게 차이 나지 않는 것으로 나타났다. 그러나 변형회귀모형이 물동량의 최근 추세를 적극 반영하거나 예견되는 예측치에 부합할 수 있도록 변수의 변형 혹은 관계함수의 다양한 변환을 통해 예측의 적합도를 향상시킨 점에 유의할 필요가 있다. 따라서 변형된 회귀모형에 의한 예측치의 MAE는 변형되기 전 회귀모형의 MAE에 비해 전반적으로 낮은 상태를 유지하고 있고 신경망모형과의 차이도 줄어들어 MAE의 상대적인 비율은 ‘상승 지속형’이 1.28에서 1.15로, ‘상승 둔화형’은 0.66에서 1.00으로 변화되었다.

각 추세유형별 적합도의 우열관계는 전반적으로 앞서의 경우와 크게 다를 바 없으나 ‘상승 둔화형’에 속하는 수출 플라스틱·고무 품목은 적합모형이 신경망모형에서 회귀모형으로 우열관계가 바뀌었고, ‘정체형’에서는 수출 기타잡화 품목을 제외한 나머지 두 품목에서 적합모형이 달라진 것으로 나타났다. 적합모형에 변화가 발생한 품목이 모두 신경망모형이 우수한 것으로 판명된 ‘상승 둔화형’과 ‘정체형’에서 나타난 것은 회귀모형의 변환을 통해 기존 회귀모형의 단점이 어느 정도 보완되었다는 점을 시사해 준다. 즉 기본적인 회귀모형의 작

위적 변형을 통해 변수 간의 관계를 선형적 함수형태로만 파악하려는 회귀모형의 단점을 보완한 것이 유효하게 작용한 것으로 판단되었다. <표 5-3>은 <표 5-1>에서 제시된 변수들 간의 관계를 변형회귀모형을 통해 나타낸 것으로 2004년 당 원의 용역연구보고서인 「전국 항만물동량 예측」에서 사용되었던 물동량 예측모형들이다.

〈표 5-3〉 변형회귀모형

종속변수	회귀 모형	R^2
y_1	$y_1 = -314.62 + 2.95x_{13} + 167.44 \ln x_{14} - 0.29x_{15}$ (-5.59) (14.72) (6.32) (-0.40)	0.998
y_2	$y_2 = -453.31 + 39.16 \ln x_1 + 37.22 \ln x_2 - 12.11x_{16}$ (-1.63) (1.58) (0.72) (-0.41)	0.974
y_3	$y_3 = 5.21 - 0.01x_2 + 0.06x_5 + 0.03x_8$ (1.07) (-1.34) (1.16) (3.23)	0.667
y_4	$y_4 = 1.14 - 0.01x_2 + 0.05x_5 + 0.05x_8$ (0.19) (-0.80) (0.81) (3.83)	0.987
y_5	$y_5 = -0.86 - 0.0007x_2 + 0.04x_5 + 0.015x_8$ (-0.29) (-0.17) (1.39) (2.53)	0.908
y_6	$y_6 = -3.11 - 0.03x_2 + 4.57x_7$ (-0.26) (-1.88) (14.40)	0.936
y_7	$y_7 = -224.05 + 39.32 \ln x_1 - 0.86 \ln x_2$ (-2.92) (11.04) (-0.06)	0.954
y_8	$y_8 = 25.94 + 4.01 \ln x_1 - 6.92 \ln x_2$ (2.02) (6.74) (-3.15)	0.964
y_9	$y_9 = -94.69 + 25.92 \ln x_1 - 5.96 \ln x_2$ (-0.91) (5.40) (-0.34)	0.610
y_{10}	$y_{10} = 73.83 + 0.72x_1/x_3 - 0.07x_2$ (1.87) (2.17) (-1.26)	0.119
y_{11}	$y_{11} = -2.58 + 0.26x_6 + 0.999 \ln x_9$ (-0.07) (1.10) (0.08)	0.789
y_{12}	$y_{12} = -38.05 - 0.003x_2 + 0.69x_4 - 0.03x_8$ (-1.68) (-0.43) (1.92) (-1.00)	0.929
y_{13}	$y_{13} = -67.01 - 0.002x_2 + 1.13x_4 - 0.02x_8$ (-1.43) (-0.17) (1.53) (-0.27)	0.963
y_{14}	$y_{14} = -100.04 + 0.0001x_2 + 1.68x_4 - 0.05x_8$ (-1.63) (0.004) (1.74) (-0.63)	0.985
y_{15}	$y_{15} = -81.74 + 0.04x_2 + 0.74x_4 + 0.17x_8$ (-1.09) (1.60) (0.68) (1.62)	0.975
y_{16}	$y_{16} = 30.90 - 6.10x_7 + 7.24x_{10}$ (6.37) (-8.41) (12.20)	0.997
y_{17}	$y_{17} = -132.80 + 0.01x_2 - 5.71x_6 - 0.05x_{11} + 2.71x_{12}$ (-6.39) (0.57) (-2.26) (-0.23) (4.24)	0.896
y_{18}	$y_{18} = -102.48 + 29.90 \ln x_{11}$ (-7.12) (9.26)	0.985
y_{19}	$y_{19} = 913.74 - 190.15 \ln x_1 + 30.48 \ln x_2 + 39.25x_6$ (1.58) (-2.77) (0.88) (3.86)	0.968
y_{20}	$y_{20} = 130.98 - 0.95x_1/x_3 + 0.03x_2$ (2.44) (-2.10) (0.45)	0.640

주 : 1) 괄호 안의 수치는 t-값임.

〈표 5-4〉 예측의 적합도 비교분석 II(변형 Regression과 Neural Network)

주세 유형	품목	MAE		R-square		MAE의 상대비율 ¹⁾	적합모델
		Regression	Nueral Network	Regression	Nueral Network		
상승 지속형	수입_유연탄	23.40	93.29	0.998	0.977	3.99	Regression
	수입_기타광석	19.65	27.39	0.974	0.952	1.39	Regression
	수입_기계류	5.79	24.12	0.987	0.767	4.17	Regression
	수입_철강	39.22	42.77	0.936	0.923	1.09	Regression
	수입_방직섬유	139.74	37.23	0.610	0.946	0.27	Neural Network
	수출_차량	34.62	34.98	0.975	0.946	1.01	Regression
	수출_방직섬유	59.03	108.63	0.968	0.899	1.84	Regression
	소계	321.46	368.42			1.15	
불규칙 상승	수입_비철금속	13.35	5.15	0.667	0.937	0.39	Neural Network
	수출_기계류	8.55	17.74	0.963	0.866	2.07	Regression
	수출_전기기기	6.56	18.05	0.985	0.889	2.75	Regression
	소계	28.46	40.93			1.44	
상승 둔화	수입_전기기기	4.69	5.48	0.908	0.845	1.17	Regression
	수입_화학공업	20.69	12.80	0.954	0.981	0.62	Neural Network
	수출_조제식품	3.38	3.18	0.789	0.810	0.94	Neural Network
	수출_비철금속	4.04	3.29	0.929	0.936	0.81	Neural Network
	수출_철강	6.28	24.06	0.997	0.969	3.83	Regression
	수출_화학공업	29.16	19.09	0.896	0.951	0.65	Neural Network
	수출_플라스틱고무	5.35	5.94	0.985	0.984	1.11	Regression
	소계	73.58	73.84			1.00	
정체	수입_플라스틱고무	2.02	2.98	0.964	0.894	1.47	Regression
	수입_기타	45.11	7.07	0.119	0.974	0.16	Neural Network
	수출_기타	26.20	20.80	0.640	0.866	0.79	Neural Network
	소계	73.33	30.85			0.42	

주 : 1) MAE of NN / MAE of Reg.

3) 예측 결과 비교 종합

MAE와 R^2 를 통해 살펴본 적합도 비교결과 회귀모형과 신경망모형 간에는 확연히 구분되는 특징을 보여주었다. ‘상승 지속형’에 속하는 7개 중 6개 품목에서, ‘불규칙 상승형’의 3개 중 2개 품목에서 회귀모형이 우수한 것으로 판명되었고 각 유형별 합산 수치에 있어서도 두 유형 모두 회귀모형이 적합도 면에서 신경망모형에 비해 우수한 것으로 나타났다. 그러나 ‘상승 둔화형’과 ‘정체 혹은 불규칙형’의 경우 앞의 두 유형과는 정 반대의 결과를 낳고 있다. ‘상승 둔화형’의 7개 품목 중 수입 전기기기와 수출 철강제품을 제외한 5개 품목에서 ‘정체 혹은 불규칙형’의 3개 품목 중 2개 품목에서 회귀모형보다 신경망모형이 더 적합도가 높은 것으로 판명되었다. 각 유형에 속한 모든 품목을 합산한 경우에도 모든 통계적 특성치는 신경망모형이 우수한 것으로 나타났다.

이상의 결과를 요약하면 지속 상승이던 불규칙 상승이던 물동량의 상승이 계속되고 있는 것으로 평가된 품목의 경우 선형관계를 바탕으로 한 회귀모형이 전반적으로 적합도 면에서 우수한 것으로 나타났다. 반면 상승이 둔화되거나 정체 혹은 추세를 판단할 수 없을 정도의 물동량 변화를 보이고 있는 품목은 대부분 과거 급속한 경제발전을 주도해 왔던 제조업 위주의 산업구조와 직·간접적으로 밀접한 관련이 있을 가능성이 높은 것들이다. 산업의 규모가 커지고 그 구조도 급격히 변화되고 있는 시대 흐름에 비추어 이와 같은 추세를 보이는 품목의 수는 차츰 늘어날 전망이다. 이러한 품목에 대한 물동량을 특정 함수모형을 통해 계량적인 접근을 통해 예측한다는 것은 불가능하거나 아주 어려운 작업이 될 수 있다. 이러한 품목들은 일정한 추세를 전제로 한 회귀분석 보다는 모형에서 필요한 분포의 가정을 통하여 가상의 생성 데이터를 통해 새로운 예측 상황을 만들어 가는 신경망모형이 보다 타당한 결과를 낳을 수 있다.

본 서에는 또한 이러한 추세유형별 일반론을 벗어난 품목에 대해 두 예측모형의 특성을 토대로 나름대로 그 이유를 분석하였다. 다만 본 서에 기술된 분

석내용은 일반론에 근거한 추론적 수준임에 유의해야 한다. 가장 큰 이유는 신경망모형에서는 초기 예측 과정에서부터 최종 예측 결과에 이르기까지 사용된 함수를 전혀 알 수 없어 예측 결과에 대한 정확한 해석이 불가능하기 때문이다.

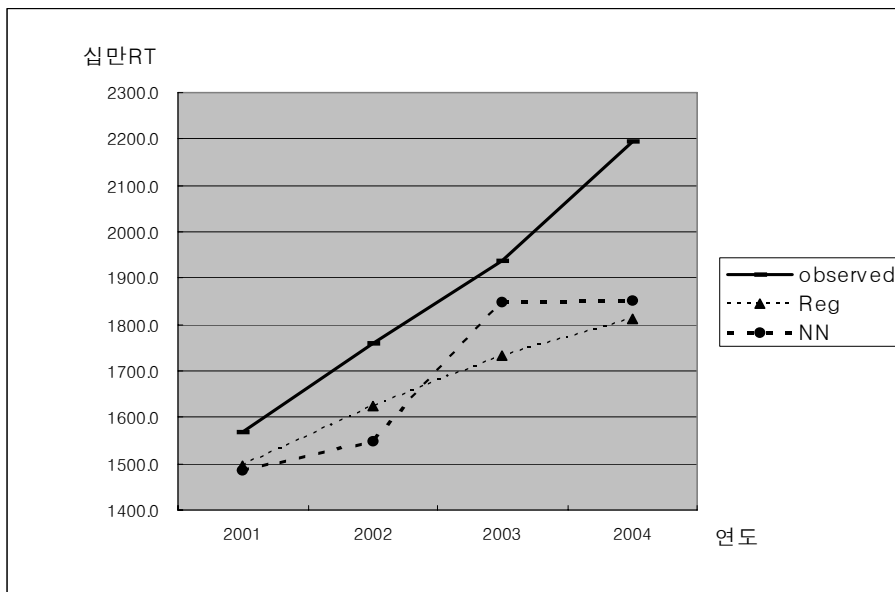
추세유형별로 일반적인 현상을 벗어난 품목의 과거 실적치(1981~2000년)의 변동추이를 비롯해 적합도의 측정 대상 연도인 2001~2004년까지의 실적치의 움직임을 자세히 분석해 본 결과 일부 품목에서 예외적인 현상이 발생한 이유를 추세유형의 구분과정 및 신경망모형과 회귀모형에 의한 예측치의 특징에서 어느 정도 찾을 수 있었다.

회귀모형은 신경망모형에 비해 모형의 수립과정에서 과거의 실적치를 고루 반영하는 반면 신경망모형은 최근의 실적을 우선시하는 특징을 보여주고 있다. ‘지속 상승형’의 변동추이를 보이는 수입방직섬유의 경우 우리나라가 겪었던 ’90년대 후반 환란위기를 넘기면서 더욱 높은 증가세를 보이고 있는데 회귀모형이 ’80년대 이후 과거의 실적을 골고루 반영하여 완만한 흐름을 보이고 있는 반면 신경망모형에서는 ’90년대 후반 이후의 높은 물동량 증가세에 크게 영향을 받고 있다. 이러한 이유로 오히려 신경망모형이 적합도에서 나은 면을 보인 것으로 평가된다. ‘불규칙 상승형’에서의 수입 비철금속과 ‘상승 둔화형’의 수입 전기기기 품목이 예외를 보인 것은 추세유형의 구분이 명확하지 않은데서 비롯된 것으로 판단된다. 수입 비철금속은 2004년의 급격한 실적 상승이 없었다면 ‘상승 둔화형’ 형태의 품목으로 간주해도 될 정도이며, 수입 전기기기는 조정기간이 환란에 따른 금융위기 이후 2000년대 초반까지로 짧다는 점과 전 고점을 역시 돌파했다는 점에서 불규칙적이나 상승형에 가까운 추세 유형에 속할 수 있다는 점이 인정된다. 즉, 두 품목 모두 추세유형 구분의 경계 부분에 위치해 있어 단기 예측기간인 2001~2004년간 물동량의 미세한 움직임에도 적합모형은 달라질 수 있을 것으로 판단된다.

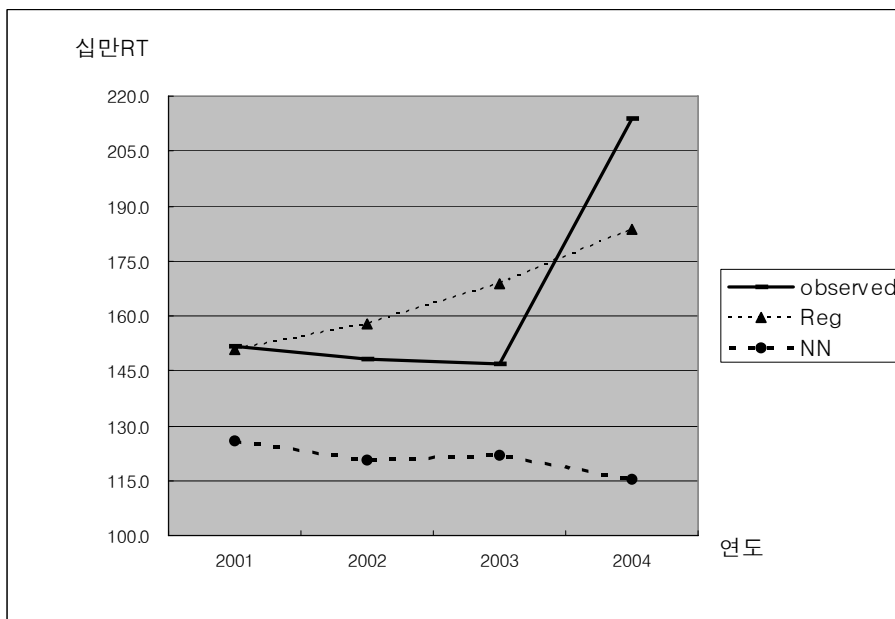
〈그림 5-8 ~ 11〉은 각 추세유형별로 품목별 물동량을 집계하여 실제치와 예측치를 비교한 것이다. ‘상승 지속형’은 회귀모형과 신경망모형 모두 실적치를 과소 추정한 경향이 있으나 모형 간 큰 차이는 보이지 않고 있다. 다만 회귀모형은 신경망모형보다는 시간이 흐를수록 물동량의 상승이 약간 높을 것으로 전망하고 있다. ‘불규칙 상승형’의 경우 회귀모형에 의한 예측치가 2004년을 제외하고는 실적치를 상회하고 있는 반면 신경망모형은 2001년 직전의 물동량 증가세 둔화에 영향 받아 실적치를 크게 과소 추정하고 있다.

‘상승 둔화형’에서는 회귀모형에 의한 예측치가 실적치를 크게 상회하고 있는 것과는 대조적으로 신경망모형은 실적치를 하회하고 있으나 실적치와 예측치와의 오차(discrepancies) 폭은 상대적으로 적은 편이다. ‘정체 혹은 불규칙형’에서는 큰 의미를 부여하긴 힘들으나 회귀모형에 의한 예측보다는 신경망모형이 훨씬 예측력이 높은 것을 알 수 있다.

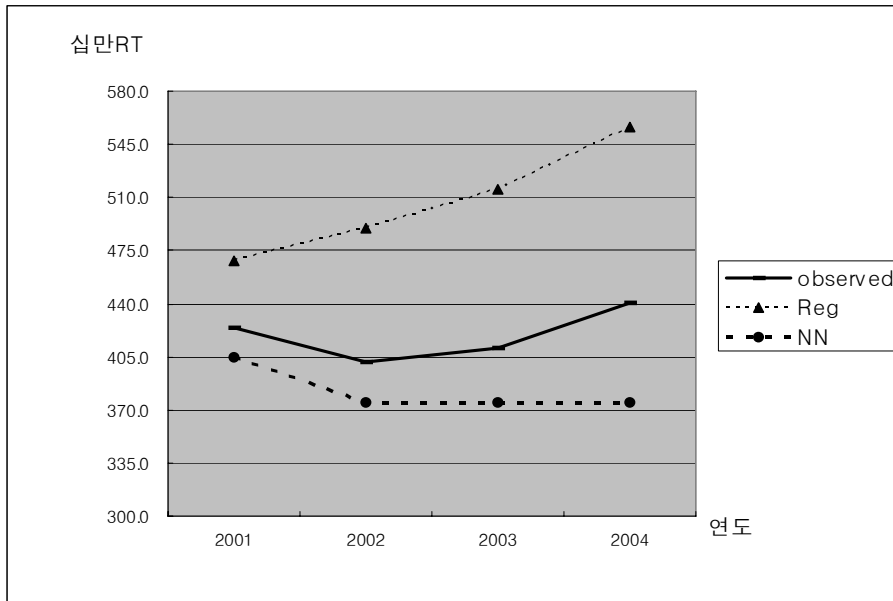
〈그림 5-8〉 ‘상승 지속형’ 품목의 2001~2004 실제치와 예측치의 비교



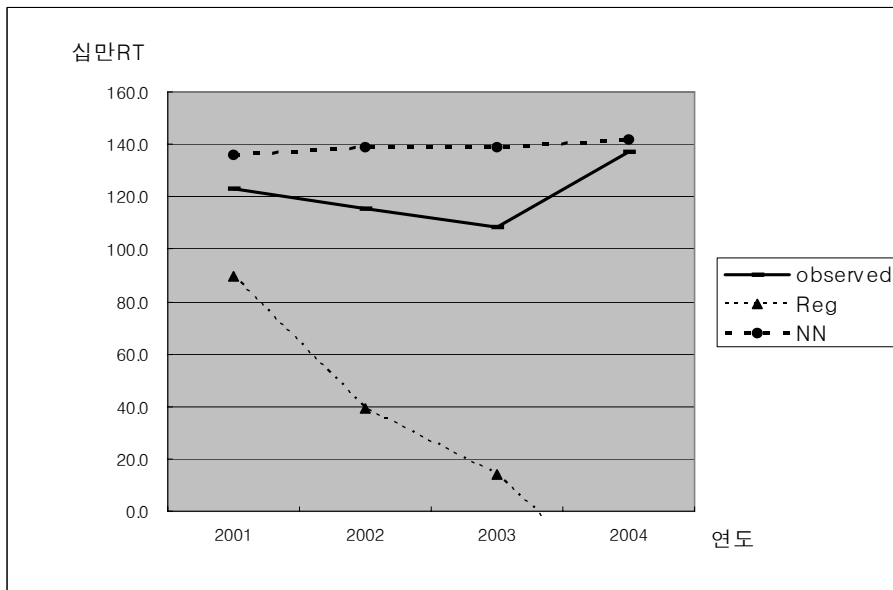
〈그림 5-9〉 ‘불규칙 상승형’ 품목의 2001~2004 실제치와 예측치의 비교



〈그림 5-10〉 ‘상승 둔화형’ 품목의 2001~2004 실제치와 예측치의 비교



〈그림 5-11〉 ‘정체 혹은 불규칙형’ 품목의 2001~2004 실제치와 예측치의 비교



제6장

몬테카를로 시뮬레이션 분석

1. 방법론

본 장은 4장에서 개발된 회귀분석모형 및 신경망 분석모형의 신뢰도 및 강건성을 측정하여 비교함으로써 두 기법의 유용성을 평가하고자 한다. 이는 몬테카를로 시뮬레이션을 이용하여 가상의 자료를 생성하고 생성된 자료의 예측값에 대한 변동성을 측정함으로써 신뢰성과 강건성을 평가하는 것을 목적으로 한다.

시뮬레이션은 회귀분석모형과 신경망분석모형을 대상으로 실시되었으며, 모형의 예측의 신뢰도 및 강건성은 20개 품목별 및 4개로 분류된 유형별로 비교되었다. 시뮬레이션 분석 결과에 대한 비교 및 판단 기준으로는 몬테카를로 방식에 의하여 생성된 각 연도별 예측치의 분포의 신뢰구간 길이 및 변동량, 그리고 극단값에 대한 민감도를 사용하였다.

2. 시뮬레이션 설계

시뮬레이션은 다음의 단계에 따라 시행되었다.

step1. 시간 t 가 경과함에 따른 오차 항 분산의 증가를 위해 다음의 오차항

을 만든다.

$$e_t = e_{t-1} + a_t, \quad a_t \sim N(0, \sigma_X^2), \quad \sigma_X^2 = \frac{1}{10} \sum_{t=t_1}^{t_2} (X_t - \bar{X})^2$$

step2. 설명변수 X 에 오차 항을 더해서 새로운 설명변수 X^* 를 만든다.

$$X_t^* = X_t + e_t$$

step3. 오차항을 추가한 새로운 설명변수를 모형에 적합시켜 예측값 \hat{Y} 를 얻는다.

$$\hat{Y}_t = \begin{cases} f_{reg}(X_t^*) & \text{regression model} \\ f_{nn}(X_t^*) & \text{Neural-Network model} \end{cases}$$

step4. 예측값 \hat{Y} 에 다시 오차 항을 추가하여 최종 예측값 \hat{Y}^* 를 얻는다.

$$\hat{Y}_t^* = \hat{Y}_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, MSE),$$

이때, MSE 는 step3의 모형에서 얻어진 평균제곱오차이다.

step5. step1~step4의 단계를 100번 반복하여 평균 추정치 $\hat{\theta}$ 를 얻는다.

$$\hat{\theta}_t = \frac{1}{100} \sum_{i=1}^{100} \hat{Y}_{ti}^*, \quad i=1, 2, \dots, 100$$

이때, \hat{Y}_{ti}^* 는 t 시점의 \hat{Y}_t^* 의 i 번째 추정치이다.

step6. step5를 1,000번 반복하여 1000개의 $\hat{\theta}_j$ 를 얻는다.

$$(j=1, 2, \dots, 1000)$$

위 과정에 의해서 얻어진 $\hat{\theta}_j$ 에서 975번째 순서통계량 $\hat{\theta}_{[.975]}$ 를 t 점에서 95% 몬테카를로 신뢰구간의 상한, 25번째 순서통계량 $\hat{\theta}_{[.25]}$ 을 t 시점에서 95% 몬테카를로 신뢰구간의 하한으로 한다.

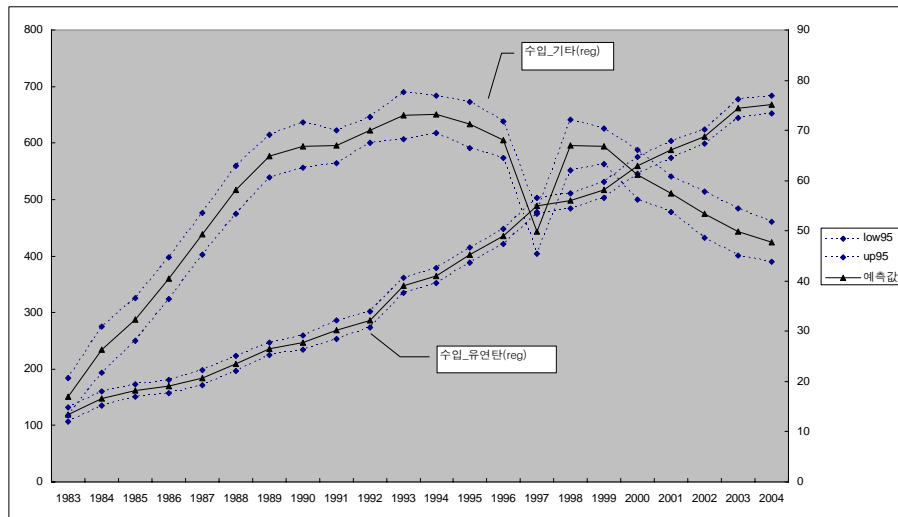
위의 6단계 시뮬레이션을 20개 품목에 대해 회귀분석과 신경망분석모형을

적용하여 보고 그 결과에서 각각 신뢰성과 강건성의 측도를 추정하여 비교하였다.

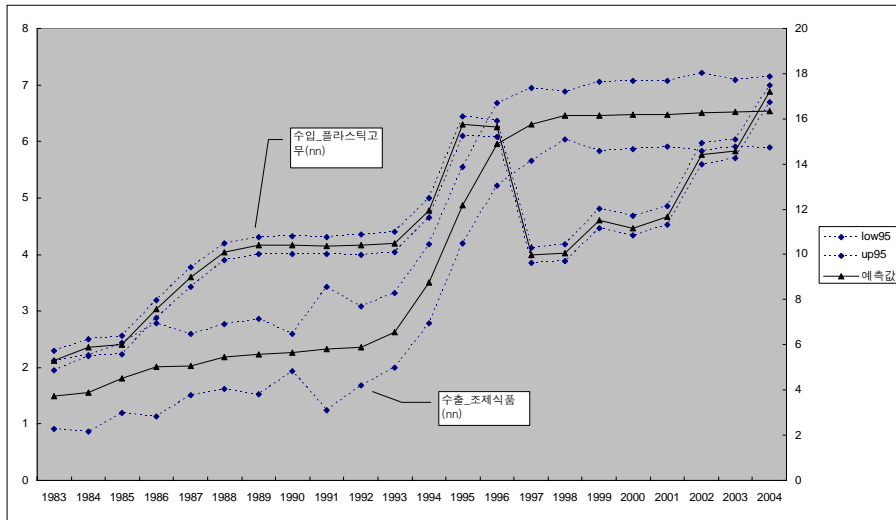
〈그림 6-1〉과 〈그림 6-2〉는 시뮬레이션의 결과로 얻어진 모형별 품목의 예측값과 신뢰구간이다.

그림에서 ‘예측값’은 오차항을 추가하지 않은 설명변수 X_t 를 이용하여 모형에 적합한 결과이고 예측값을 둘러싸고 있는 up95와 low95는 모형의 95% 몬테카를로 신뢰구간을 말한다. 신뢰구간은 시간에 따라 점차 증가하거나 또는 변화 없이 유지되는 등 품목 및 모형에 따라 다양한 형태를 보이고 있으며, 이 신뢰구간(또는 표준편차)의 길이와 추이를 통해 신뢰성과 강건성을 비교해 볼 수 있다.

〈그림 6-1〉 수입_기타와 수입_유연탄 품목의 회귀분석모형 예측값과 95% 몬테카를로 신뢰구간



〈그림 6-2〉 수입_플라스틱고무와 수출_조제식품 품목의 신경망분석모형 예측값과 95% 몬테카를로 신뢰구간



3. 신뢰성 비교

신뢰성의 측도는 각 시점별 추정치 $\hat{\theta}_j$ 의 표준편차의 합을 이용하였다. $\hat{\theta}_j$ 는 j 번째 시행에서 해당품목의 평균추정치이므로 이 추정치의 표준편차가 작다는 것은 예측의 신뢰성이 높다는 뜻으로 해석할 수 있다. 따라서 각 시점별 추정치의 표준편차 S_j 를 구한 후 각 시점에서 S_j 가 작은 모형이 더 신뢰할 수 있는 모형이 되므로 S_j 의 합을 각 모형의 신뢰성을 비교하는 측도로 보았다.

신뢰성 비교는 1983~2004년의 데이터를 이용하였다. 2005년 이후의 예측값은 과거 데이터를 이용하여 만든 모형의 적합도가 떨어지므로 신뢰성을 비교하는 데 적절하지 않다.

신뢰성 측도를 계산하는 방법은 다음과 같다.

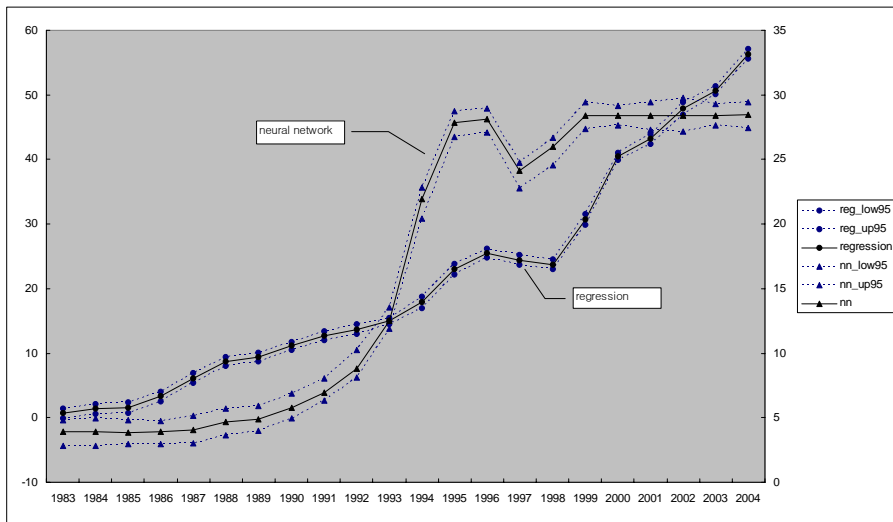
$$S = \sum_{t=1983}^{2004} S_t, \quad S_t = \sqrt{\frac{1}{999} \sum_{j=1}^{1000} (\hat{\theta}_{tj} - \hat{\mu}_t)^2}$$

여기서 $\hat{\mu}_t = \frac{1}{1000} \sum_{j=1}^{1000} \hat{\theta}_{tj}$ 이다.

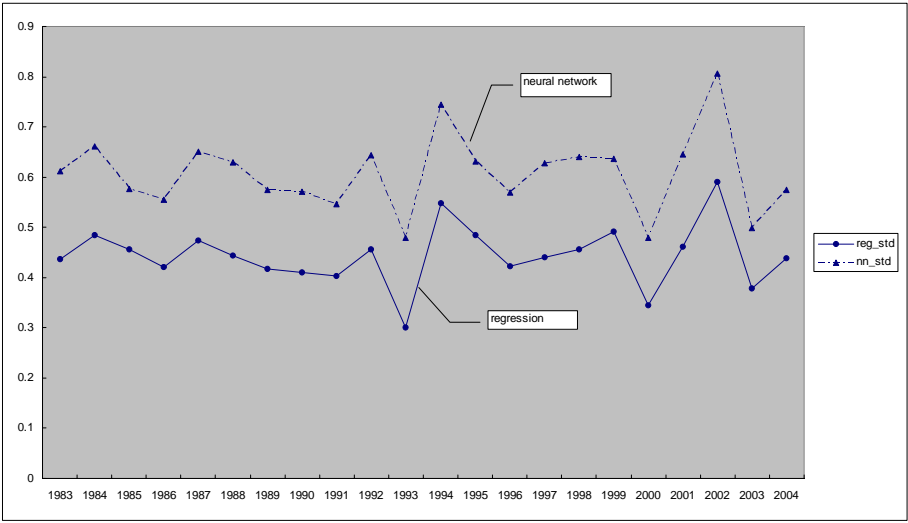
〈그림 6-3〉과 〈그림 6-5〉, 〈그림 6-7〉은 모형별 품목의 예측값과 몬테카를로 신뢰구간을, 〈그림 6-4〉와 〈그림 6-6〉, 〈그림 6-8〉은 표준편차 S_t 를 비교한 것이다.

〈그림 6-4〉, 〈그림 6-6〉, 〈그림 6-8〉의 표준편차 그래프는 각 모형별 추정치 분산의 변화추이를 확인할 수 있으며 각 그래프의 아래쪽 면적이 작은 모형이 더 신뢰성이 높다고 판단할 수 있다.

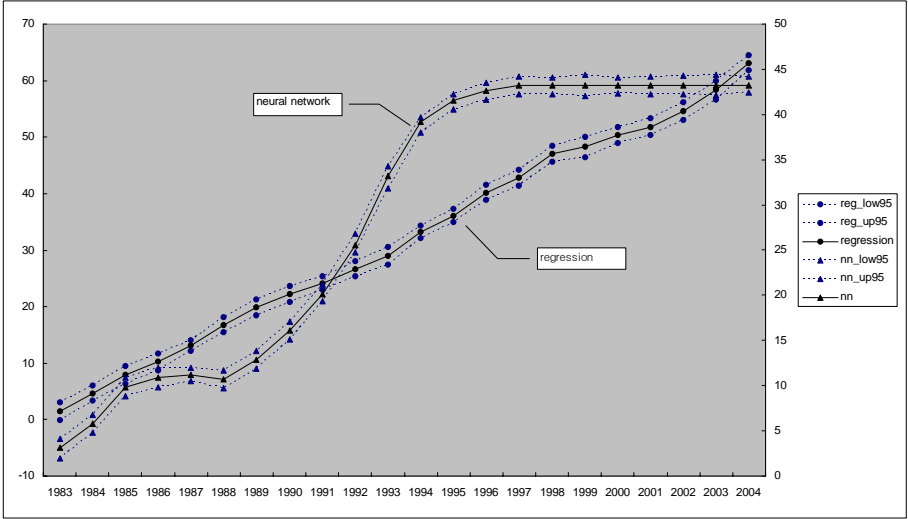
〈그림 6-3〉 수입_기계류 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 비교



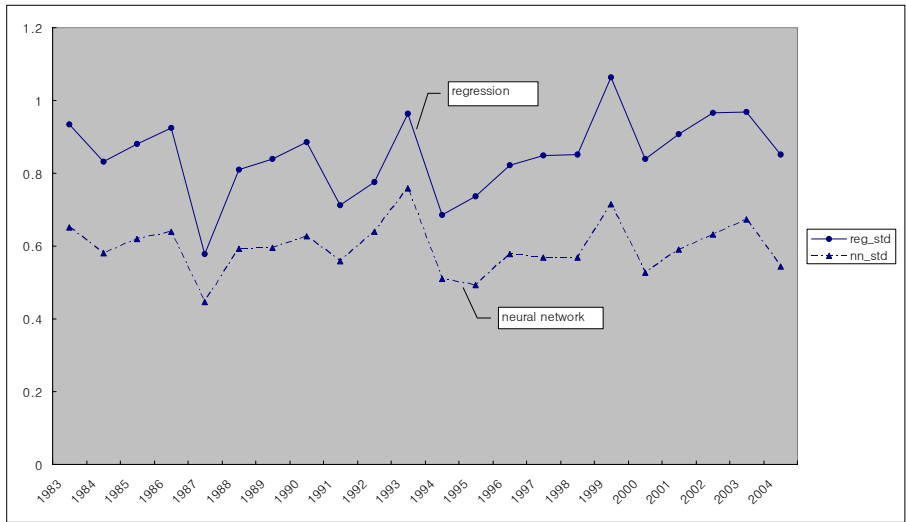
〈그림 6-4〉 수입_기계류 품목의 회귀분석모형과 신경망분석모형의 표준편차 비교



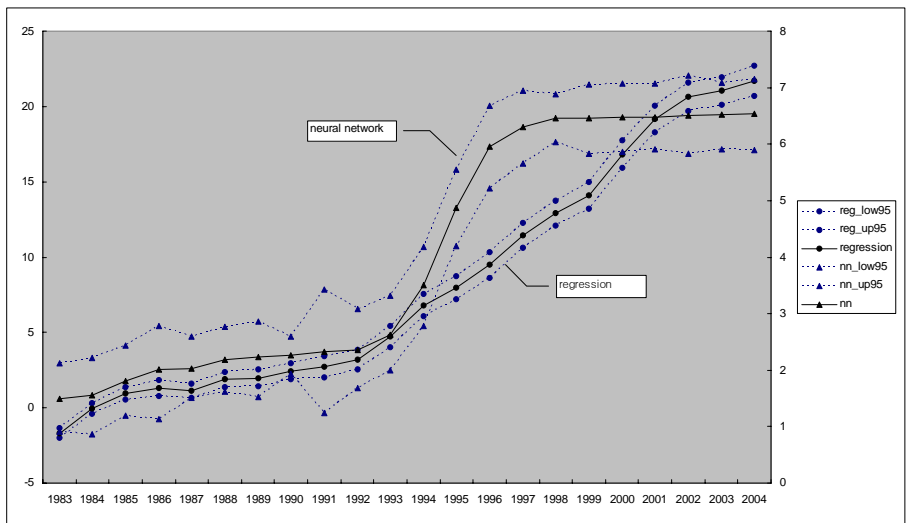
〈그림 6-5〉 수출_기계류 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 비교



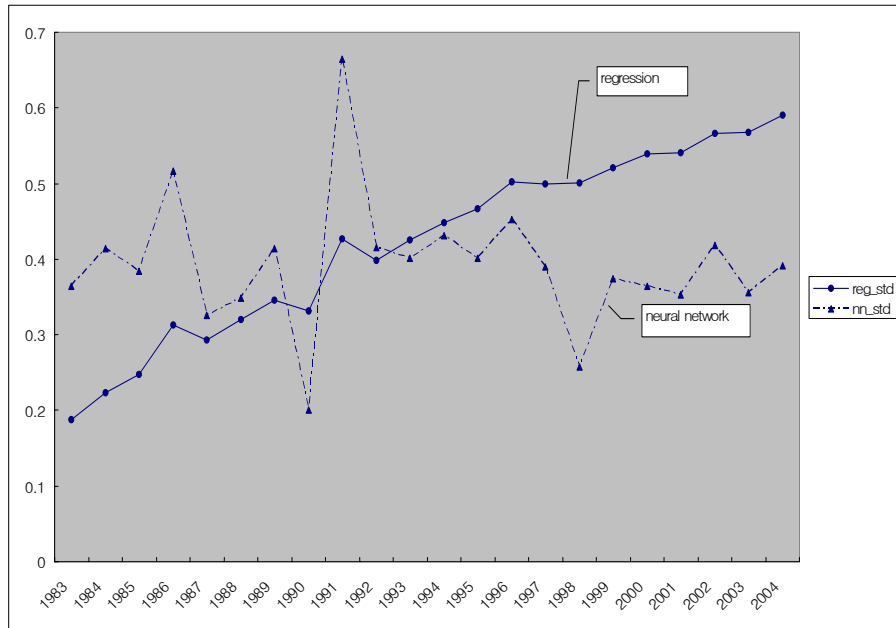
〈그림 6-6〉 수출_기계류 품목의 회귀분석모형과 신경망분석모형의 표준편차 비교



〈그림 6-7〉 수출_조제식품 품목의 회귀분석모형과 신경망분석모형의
신뢰구간 비교



〈그림 6-8〉 수출_조제식품 품목의 회귀분석모형과 신경망분석모형의
표준편차 비교



표준편차의 합 S 를 이용한 각 품목의 모형별 신뢰성 비교결과는 〈표 6-1〉에 정리되어 있으며 유형별로는 상승 지속형의 경우 회귀분석이, 불규칙 상승형의 경우는 신경망분석이 더 신뢰성이 높다고 할 수 있다. 상승 둔화형과 정체형 역시 신경망분석의 신뢰성이 더 높게 나타났다.

〈표 6-1〉 모형별 신뢰성 비교분석(Regression과 Neural Network)

추세 유형	품목	S		적합모델
		Regression	Nueral Network	
상승 지속형	수입_유연탄	180.64	204.24	Regression
	수입_기타광석	35.13	3.76	Neural Network
	수입_기계류	9.76	13.35	Regression
	수입_철강	20.78	29.37	Regression
	수입_방직섬유	23.61	47.84	Regression
	수출_차량	30.19	20.62	Neural Network
	수출_방직섬유	29.05	4.93	Neural Network
불규칙 상승	수입_비철금속	7.88	3.67	Neural Network
	수출_기계류	18.67	13.10	Neural Network
	수출_전기기기	24.96	17.30	Neural Network
상승 둔화	수입_전기기기	4.67	2.19	Neural Network
	수입_화학공업	19.26	17.28	Neural Network
	수출_조제식품	9.25	8.64	Neural Network
	수출_비철금속	8.99	4.11	Neural Network
	수출_철강	15.81	33.37	Regression
	수출_화학공업	16.52	10.34	Neural Network
	수출_플라스틱고무	11.06	7.97	Neural Network
정체	수입_플라스틱고무	3.52	5.42	Regression
	수입_기타	55.22	27.08	Neural Network
	수출_기타	79.20	27.22	Neural Network

4. 강건성 비교

강건성은 모형이 얼마나 일정한 형태를 유지할 수 있는가의 문제이다. 즉, 설명변수에 어떤 사건 등의 외부충격(shock)으로 인해 예년과 다른 극단적인 값이 발생하였을 때, 모형이 이에 크게 영향을 받지 않고 일정한 패턴을 유지할 수 있는지를 판단하는 측도를 말한다. 이는 이상치에 의한 몬테카를로 신뢰구간의 길이변화로 확인할 수 있을 것이다. 모형이 이상치에 영향을 많이 받는다면 t^* 시점의 평균추정치 $\hat{\theta}_{t^*}$ 의 신뢰구간은 영향을 받는 만큼 확장될 것이다. 이러한 성질을 확인하기 위해서는 특정 시점 t^* 에 가상의 이상치를 투여한 후 이 때 신뢰구간의 변화량을 측정하고 각 모형별 변화량을 비교함으로써 판단할 수 있다.

강건성 비교를 위한 실험 설계는 기존의 시뮬레이션 설계와 거의 흡사하나 설명변수의 특정 시점에 이상치를 투여하는 과정만 차이가 있다. 강건성의 실험설계는 다음과 같다.

step1. 시간 t 가 경과함에 따른 오차 항 분산의 증가를 위해 다음의 오차 항을 만든다.

$$e_t = e_{t-1} + a_t, \quad a_t \sim N(0, \sigma_X^2),$$

$$\sigma_X^2 = \frac{1}{10} \sum_{t=t_1}^{t_2} (X_t - \bar{X})^2, \quad t_1 = 2005\text{년}, \quad t_2 = 2015\text{년}$$

step2. 설명변수 X 에 오차 항을 더해서 새로운 설명변수 X^* 를 만든다.

$$X_t^* = X_t + e_t$$

step3. 이상치를 발생시키기 위해 특정시점 t^* 의 임의의 설명변수 $X_{t^*}^*$ 를 다음과 같이 조정한다.

$$X_{t^*}^* \leftarrow X_{t^*}^* \times 1.5$$

step4. 오차항과 이상치를 발생시킨 새로운 설명변수를 모형에 적합시켜 예측값 \hat{Y} 을 얻는다.

$$\hat{Y}_t = \begin{cases} f_{reg}(X_t^*) & \text{regression model} \\ f_{nn}(X_t^*) & \text{Neural-Neural model} \end{cases}$$

step5. 예측값 \hat{Y} 에 다시 오차 항을 추가하여 최종 예측값 \hat{Y}^* 를 얻는다.

$$\hat{Y}_t^* = \hat{Y}_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, MSE),$$

이때, MSE 는 step4의 모형에서 얻어진 평균제곱오차이다.

step6. step1~step5의 단계를 100번 반복하여 평균 추정치 $\hat{\theta}$ 를 얻는다.

$$\hat{\theta}_t = \frac{1}{100} \sum_{i=1}^{100} \hat{Y}_{ti}^*, \quad i=1, 2, \dots, 100$$

이때, \hat{Y}_{ti}^* 는 t 시점의 \hat{Y}_t^* 의 1번째 추정치이다.

step7. step6를 1000번 반복하여 1000개의 $\hat{\theta}_j$ 를 얻는다.

$$(j=1, 2, \dots, 1000)$$

위 과정에 의해서 얻어진 $\hat{\theta}_j$ 에서 975번째 순서통계량 $\hat{\theta}_{[.975]}$ 를 t 시점에서 95% 몬테카를로 신뢰구간의 상한, 25번째 순서통계량 $\hat{\theta}_{[.25]}$ 을 t 시점에서 95% 몬테카를로 신뢰구간의 하한으로 한다.

특정 시점 t^* 에서 임의의 설명변수에 인위적인 이상치를 발생시켰으므로 $\hat{\theta}_{t^*}$ 는 다른 시점에 비해 보다 넓은 신뢰구간을 갖게 되며, 이 때 t^*-1 시점과 t^* 시점의 신뢰구간의 변화량을 파악하면 이상치에 의한 효과를 측정할 수 있다.

이를 수식으로 정리하면 다음과 같다.

t 시점에서 95% 신뢰구간의 길이를 D_t 라고 하면

$$D_t = |\hat{\theta}_{[.975]} - \hat{\theta}_{[.25]}|$$

이고, 시점 t^* 의 신뢰구간의 길이는 D_{t^*} 이다. 이 때, 바로 전 시점 t^*-1 에서의 신뢰구간의 길이 D_{t^*-1} 과 D_{t^*} 의 차이가 바로 이상치로 인한 효과이다.

즉, 이상치 효과 E_{t^*} 는

$$E_{t^*} = \frac{|D_{t^*} - D_{t^*-1}|}{D_{t^*-1}} \times 100$$

가 된다.

강건성 비교 역시 1983년부터 2004년까지의 설명변수 예측값을 이용하여 분석하였다. 외부충격이 발생하지 않은 상황에서 설정된 모형이 만약 외부충격이 발생했다면 얼마나 크게 반응하는지 확인하는 과정이므로 강건성 분석으로 합리적인 방법이다.

강건성 비교를 위한 시뮬레이션 상에서 이상치 발생시점 t^* 는 1995년으로 하였다.

〈그림 6-9〉와 〈그림 6-11〉, 〈그림 6-13〉은 이상치를 넣었을 때 모형의 예측값과 신뢰구간을 표시한 것이고 〈그림 6-10〉과 〈그림 6-12〉, 〈그림 6-14〉는 모형별 신뢰구간의 길이 D_t 를 표시한 것이다.

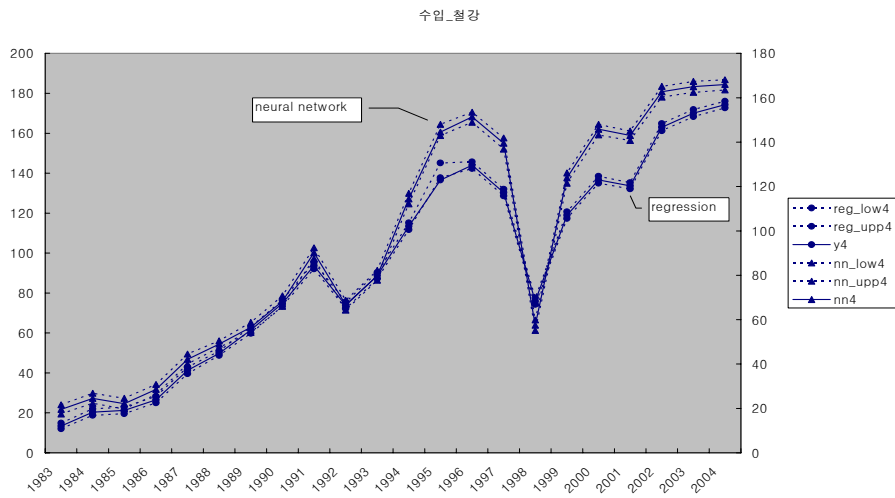
외부충격에 의한 신뢰구간 변화량은 수입_철강과 수출_기계 품목에서는 신경망 분석이 더 적은 변화를 보이고 있고 수출_철강품목에서는 회귀분석이 더 적은 것으로 보인다. 〈그림 6-10〉, 〈그림 6-12〉, 〈그림 6-14〉는 이러한 상황을 잘 보여주고 있다.

〈표 6-2〉 모형별 강건성 비교를 보면 상승 지속형에서 회귀분석이 더 강건한 방법이지만, 불규칙 상승형, 상승둔화형, 정체형에서는 신경망 분석이 더 강건성이 높다는 것을 확인할 수 있다.

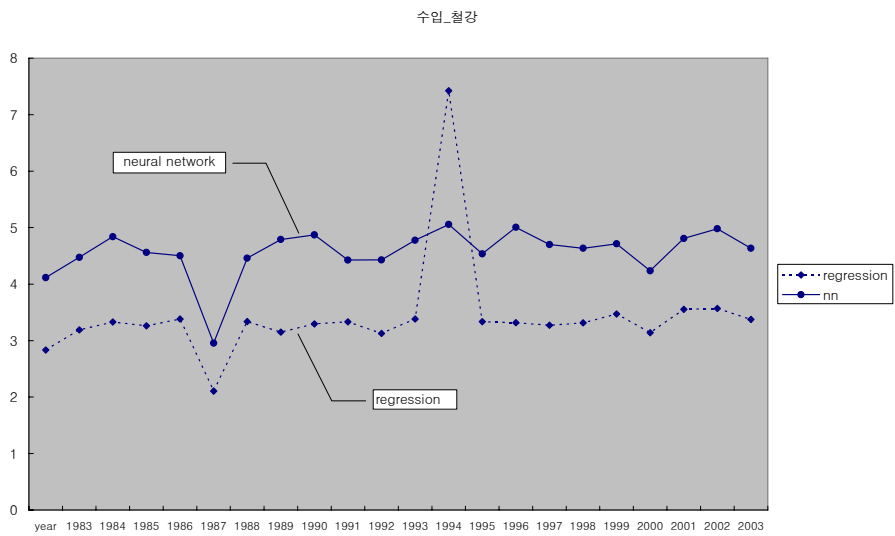
〈표 6-2〉 모형별 강건성 비교분석(Regression과 Neural Network)

추세 유형	품목	E_{1995}		적합모델
		Regression	Neural Network	
상승 지속형	수입_유연탄	18.57%	32.60%	Regression
	수입_기타광석	35.68%	39.07%	Regression
	수입_기계류	9.33%	19.84%	Regression
	수입_철강	119.55%	5.87%	Neural Network
	수입_방직섬유	23.89%	33.99%	Regression
	수출_차량	39.27%	54.54%	Regression
	수출_방직섬유	101.89%	195.11%	Regression
불규칙 상승	수입_비철금속	94.94%	13.63%	Neural Network
	수출_기계류	82.43%	0.27%	Neural Network
	수출_전기기기	65.19%	200.18%	Regression
상승 둔화	수입_전기기기	74.48%	173.24%	Regression
	수입_화학공업	57.86%	35.80%	Neural Network
	수출_조제식품	19.88%	3.66%	Neural Network
	수출_비철금속	111.13%	21.11%	Neural Network
	수출_철강	128.82%	785.38%	Regression
	수출_화학공업	233.72%	126.78%	Neural Network
	수출_플라스틱고무	88.16%	17.67%	Neural Network
정체	수입_플라스틱고무	7.89%	6.12%	Neural Network
	수입_기타	355.69%	28.87%	Neural Network
	수출_기타	437.90%	108.80%	Neural Network

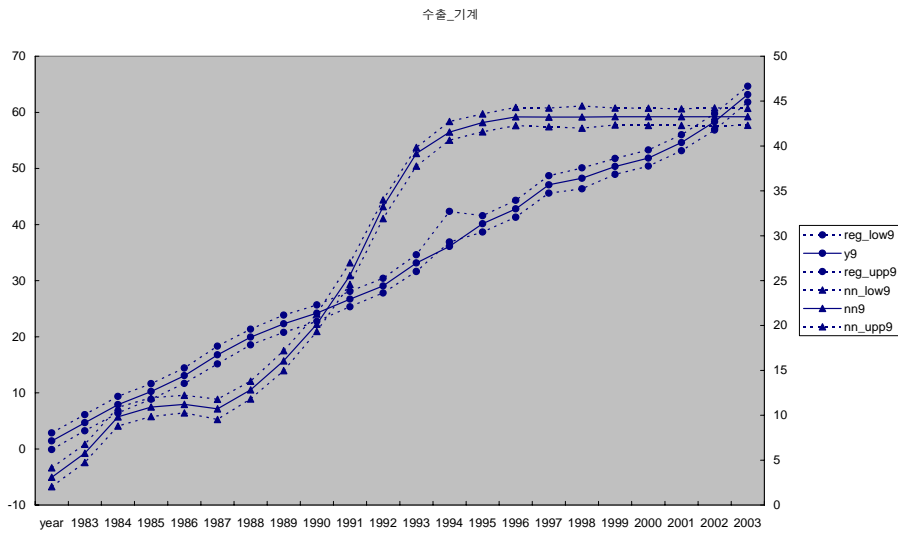
〈그림 6-9〉 수입_철강 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 비교



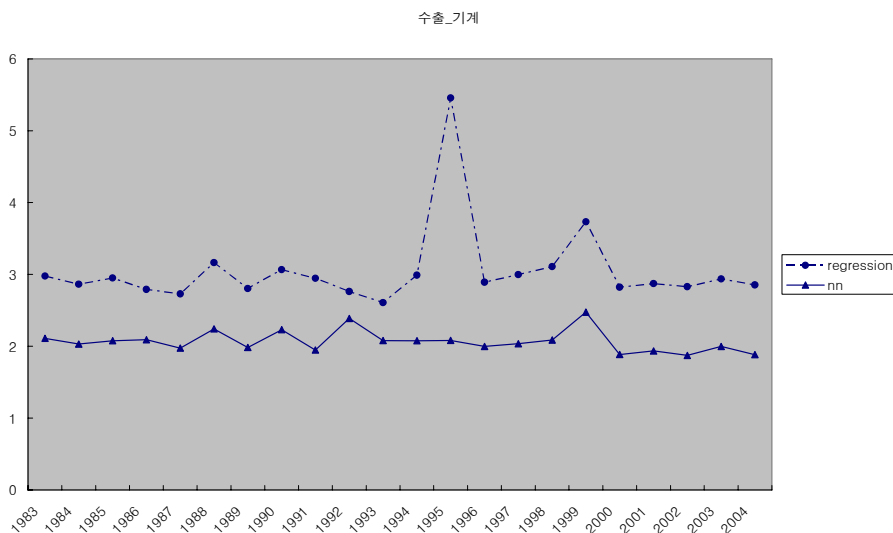
〈그림 6-10〉 수입_철강 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 길이비교



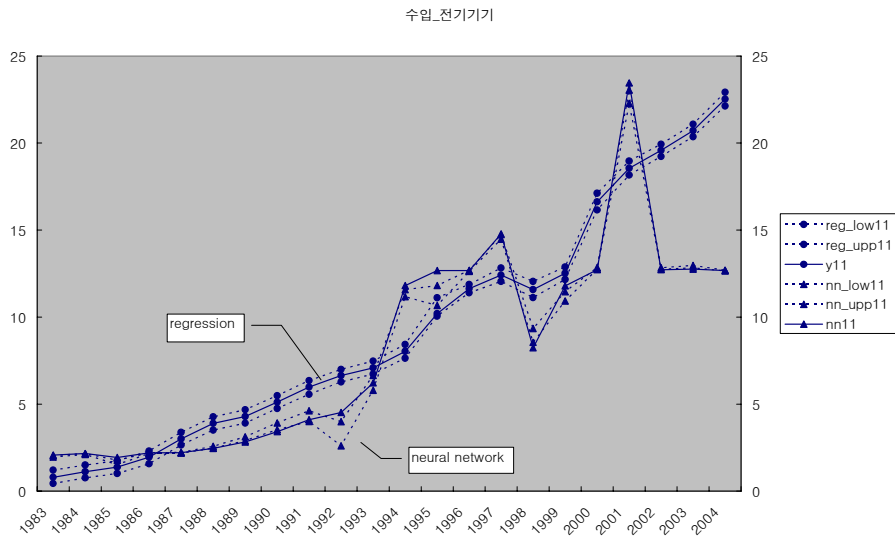
〈그림 6-11〉 수출_기계 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 비교



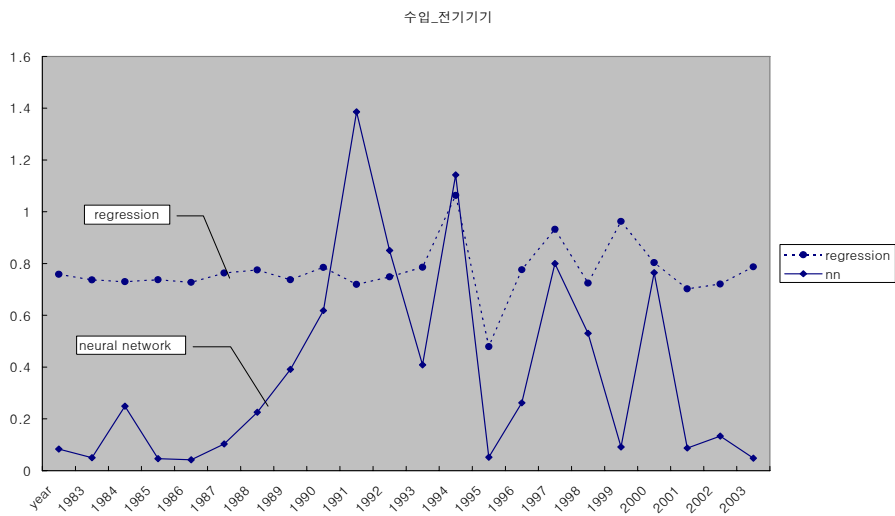
〈그림 6-12〉 수출_기계 품목의 회귀분석모형과 신경망분석모형의
신뢰구간 길이비교



〈그림 6-13〉 수출_철강 품목의 회귀분석모형과 신경망분석모형의 신뢰구간 비교



〈그림 6-14〉 수출_철강 품목의 회귀분석모형과 신경망분석모형의
신뢰구간 길이비교



5. 비교 종합

변동성을 통해 살펴본 신뢰도 비교결과 회귀모형과 신경망모형 간에는 서로 다른 특징을 지니고 있었다. 회귀분석은 상승 지속형인 경우에 신뢰성이 우수한 것으로 볼 수 있으나 신경망 분석은 불규칙 상승형, 상승둔화형, 정체형에서 더 우수한 것으로 나타났다.

강건성의 경우에도 상승 지속형에서 회귀분석이 더 강건한 방법이지만, 불규칙 상승형, 상승둔화형, 정체형에서는 신경망 분석이 더 강건한 방법인 것으로 확인되었다.

제7장

결론 및 정책대안

1. 결론

추세유형별 항만물동량 예측시 회귀모형과 신경망모형을 비교·분석해본 결과 상승 지속형의 경우 적합성, 신뢰성, 강건성 등 모든 통계적 특성에서 회귀모형이 우수한 것으로 판명되었다. 반면 상승 둔화형과 정체형은 3가지 측면에서 모두 신경망모형이 더 나은 것으로 나타났다. 불규칙 상승의 경우 적합성 측면에서는 회귀모형이, 신뢰성과 강건성 측면에서는 신경망모형이 더 좋은 것으로 나타나 특성치별로 차이를 보였다.

특히 본 연구는 기존 연구에서 시도하지 않았던 강건성에 대한 모형별 우위성을 시뮬레이션을 통해 판단하였다. 신경망모형과 타 통계기법과의 효능 비교 시 선행연구에서 강건성에 높은 비중을 두지 않은 것은 신경망모형 특유의 성질 때문인 것으로 판단된다. 즉, 신경망모형이 모델을 구축할 때는 최적의 것이 구축될 때까지 새로이 생성된 데이터로 모델을 평가하기³¹⁾ 때문에 미래 예측치의 경우 기본적으로 다른 통계적 방법론보다 강건할(robust) 가능성이 높기 때문이다. 그럼에도 본 연구에서는 강건성에 대한 비교 분석결과가 다른 통계적 특성치 비교 결과와 크게 유의하지 않은 것으로 나타났다.

본 연구를 종합해 보면 지속 상승이든 불규칙 상승이든 물동량의 상승이 계

31) 강현철·한상태·최종후·김은식·김미경, 「SAS Enterprise Miner 4.0을 이용한 데이터 마이닝」, 2001. 12.

속되고 있는 것으로 평가된 품목의 경우 선형관계를 바탕으로 한 회귀모형이 전반적으로 적합도 면에서 우수한 것으로 나타났다. 반면 상승이 둔화되거나 정체 혹은 추세를 판단할 수 없을 정도의 물동량 변화를 보이고 있는 품목은 모형에서 필요한 분포의 가정을 통하여 가상의 생성 데이터를 통해 새로운 예측 상황을 만들어 가는 신경망모형이 보다 타당한 결과를 내고 있다.

이것은 홍콩 항의 항만물동량 예측에 있어 신뢰성과 정확성 면에서 신경망 모형이 회귀모형보다 전반적으로 나은 것으로 나타난 Lam, W. H. K. and Pan L. P. and Eddie C. M.(2004)³²⁾의 연구결과와, 비교적 짧은 기간 동안 등락폭이 심하게 발생하는 품목의 물동량 예측시 신경망모형의 우수성을 입증한 신성환(1995)³³⁾의 이론과도 부합하고 있다. 신성환(1995)의 이론은 평균 또는 추세선을 크게 이탈하지 않는 물동량보다 단기간에 급등과 급락을 반복하는 형태를 보이는 물동량 예측에 회귀분석이나 시계열모형보다는 신경망모형이 보다 적절하다는 정효준·이홍근(2002)³⁴⁾의 연구결과와도 합치되고 있다.

향후 산업이 고도화되고 그 구조도 급격히 변화되고 있는 시대 흐름에 비추어 이러한 품목에 대한 물동량을 예측하는 데에 신경망모형은 유용하게 사용될 수 있다. 특히 회귀분석의 경우 데이터의 정규성을 가정할 뿐 아니라, 선형성을 가정하는데 이러한 엄격한 조건을 만족하는 데이터는 현실적으로 많지 않다. 가령 데이터에 outlier가 존재하거나, 데이터의 분포 자체가 한쪽으로 치우치는 등 정규성을 따르지 않을 경우에는, 추정치의 적합성과 신뢰성에 많은 문제가 발생할 수 있다.

그러나 이러한 장점에도 불구하고 인공신경망의 가장 큰 약점은 어떠한 변수가 중요한지, 어떻게 상호작용이 이루어져 그러한 결과물을 낳게 되는지에 대한 설명은 할 수가 없다. 따라서 변수 상호간의 인과관계에 대한 설명력 보다는 정확한 예측이 중요할 경우에만 이용될 수 있다. 최근 인공신경망의 해석

32) Lam, W. H. K., and Pan L. P., and Eddie C. M, 전게서, 2004.

33) 신성환, 전게서, 1995.

34) 정효준·이홍근, 전게서, 2002.

력을 보완하기 위해 인공신경망 분석 후 의사결정나무를 사용하여 해석을 돕는 방법이 사용되고 있다. 인공신경망과 의사결정나무가 많은 통계 프로그램에 적재되고 있음은 이미 전술한 바 있다.

또 신경망모형의 불해석력을 해결하는 방안으로 선행연구에서 많은 연구자들이 회귀분석 혹은 시계열모형과 같은 비선형모형과의 통합 혹은 검용 사용을 제안하고 있다. 분야에 따라 그 구체적인 용도는 다르다고 할 수 있으나 예측 결과에 대한 변수 상호간의 인과관계의 해석은 인공신경망모형이 적극 활용되기 위해 극복해야 할 과제임에는 틀림없다.

2. 정책대안

현재 정부의 항만개발계획은 항만물동량에 대한 중장기적인 예측을 근거로 ‘전국 무역항 기본계획’이 10년마다 수립되고, 5년마다 이에 대한 계획의 수정 및 보완작업이 이루어지도록 하고 있다. 따라서 항만물동량에 대한 예측의 목표연도가 예측시점부터 10년 이상 20년까지에 초점을 맞추고 있어 예측기간이 너무 길다는 문제점을 안고 있다. 즉, 최근처럼 대내외적인 물류여건이 급변하고 있는 시기에 먼 장래에 대한 예측이 정확하기를 기대하기 어렵고, 5년마다 이루어지는 보완 및 수정시기도 변화요인을 충분히 반영하기에는 그 간격이 너무 크다는 것이다.

정부가 당 원에 ‘수요예측센터’를 설치하여 항만물동량의 변동성 감지 및 이를 실시간 예측에 반영할 수 있도록 한 것도 정부의 장기 항만개발계획을 좀 더 탄력적으로 운영하자는 뜻이 담겨 있다. 항만물동량에 영향을 미칠 수 있는 대내외적 요인의 변동성을 지속적으로 반영하고, 나아가 장기적인 항만개발계획까지 융통성 있게 수정할 수 있기 위해서는 물동량에 대한 단기적인 예측이 정기적으로 이루어질 필요가 있다. 예측은 분기별 혹은 반기별로 실시하

되 분석 자료는 월별 혹은 분기별 단위의 자료를 바탕으로 하는 것이 바람직할 것으로 판단된다.

월별 혹은 분기별 자료를 이용한 단기적인 예측은 가용한 데이터의 용량이 많아지고 장기적인 예측에 비해 예측의 정확도가 보다 중요시 된다. 따라서 물동량에 미치는 요인(factor 혹은 input variable)이나 여건이 크게 변화될 경우 이에 따른 단기적인 예측은 본 연구에서 실증된 내용과 모형 본래의 특성을 고려할 때 신경망모형이 요긴하게 사용될 수 있으며 자료가 축적될수록 그 효용성은 더욱 높아질 수 있다.

물론 물동량의 변화에 대처한 정책적인 판단에 도움이 되기 위해서는 물동량에 영향을 미치는 변수와의 인과관계 파악을 위한 정형화된 예측함수가 요구될 수 있다. 이럴 때 데이터 마이닝의 예측기법 중 하나인 의사결정나무가 보조적으로 사용될 수 있을 것이다. 그러나 물동량의 변화에 다양한 요인이 복합적으로 작용하여 나타나는 현상인 만큼 특정 요인의 변화에 따른 영향만을 별도로 추정하기는 어려운 일이다. 따라서 단기적인 예측은 그 자체의 정확도에 의미를 두는 것이 바람직할 것이다.

항만의 개발에 소요되는 시간은 대체로 5년 전후이다. 또한 항만의 국제 경쟁력 강화를 위해서는 항만개발에 대한 장기적 비전이 미리 제시되어야 한다. 이런 측면에서 항만물동량에 대한 장기적인 예측의 중요성은 과거와 달라질 수 없다. 장기적인 예측은 단기적인 것과는 달리 국가경제발전 및 정책방향과 궤를 같이 할 수밖에 없으며 따라서 이를 반영할 수 있는 구체적 함수모형의 설정이 더욱 중요시된다. 본 연구에서는 이를 위해 신경망모형의 불해석력을 해결하는 방안으로 회귀분석과의 통합 혹은 겸용 사용을 제안하고자 한다. 이미 선행연구에서 많은 연구자들이 이와 유사한 안을 제시한 바 있으나 그 체계적인 용도는 분야별로 각기 다름을 전술한 바 있다. 상술하자면 품목별 특성에 따라 신경망모형이 예측이 유리한 것으로 판단될 경우 우선 신경망모형을 이용한 예측치를 산출한다. 그 다음 회귀모형을 적절히 변형시켜 동 예측 결과에 가장 근접한 모형을 최종 예측모형으로 설정하는 것이다. 즉, 신경망모형을 통

해 얻어진 예측 결과를 모형선택의 기준으로 삼고 이에 유사한 결과를 얻도록 기본적인 회귀모형의 작위적 변형을 가하도록 하는 것이다. 이 때 작위적 모형의 설정은 설명변수와 독립변수와의 관계에 적절한 것으로 평가되는 함수적 변환만이 고려되어야 한다. 최종 예측 결과는 물론 선택된 회귀모형에서 도출된 것으로 삼는다. 이와 관련해 제5장에서는 회귀모형의 함수변환을 통해 기존 회귀모형의 단점이 어느 정도 보완될 수 있다는 점을 언급한 바 있다.

본 연구는 최근 우리나라 산업구조의 급변에 따른 물동량의 변화를 주어진 과거 자료의 추세적 변화요인을 통해서만 파악하려는 기존 예측방법의 한계를 개선할 수 있는 대안을 찾기 위해 이루어졌다. 항만물동량에 대한 신경망모형의 효용 가능성은 통계적 방법을 통해 검증하였으나, 실제 적용은 모형을 구조적으로 설명해 줄 수 있는 함수가 없는 문제점을 어떻게 해결하느냐에 달려 있다고 할 것이다.

참고문헌

- 강현철·한상태·최종후·김은석·김미경, “SAS Enterprise Miner 4.0을 이용한 데이터 마이닝”, 2001. 12
- 김경태·이장형·최현호, “신경망에 의한 주가수익률 예측모형”, 「기업경영연구」, 제14집, 2001. 5.
- 김태훈·홍한국, “회귀모형과 신경망모형을 이용한 아파트 가격 모형에 관한 연구”, 「국토연구」, 제43권, 2004. 12.
- 노대협·이택호·한인구, “인공신경망-금융시계열 모형을 이용한 KOSPI 200 주가지수의 변동성 예측”, 「경영학연구」, 제34권 제3호, 2005. 6.
- 손민우·이길성, “신경망 이론과 회귀분석을 통한 홍수위 예측”, 「대학토목공학회논문집」, 제23권 제3B호, 2003. 5.
- 신성환, “인공신경모형과 이동평균법의 환율예측력 평가”, 「금융연구」, 제9권 1호, 1995.
- 전찬영, “교차검증을 통한 우리나라 중장기 항만물동량 예측”, 「해양정책연구」, 제14권, 1999. 8.
- 정효준·이흥근, “ARIMA 모형과 인공신경망모형의 BOD예측력 비교”, 「한국환경위생학회지」, 제28권 제3호, 2002.
- 해양수산부, 「전국 항만물동량 예측」, 2004. 12.
- B. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.
- I. Witten & E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, 2000.
- L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*, Wadsworth, 1984.

- Lam, W. H. K., and Pan L. P., and Eddie C. M, “Forecasts and Reliability Analysis of Port Cargo throughput in Hong Kong” , *Journal of Urban Planning and Development*, Sep. 2004.
- M. Berry & G. Linoff, *Data Mining Techniques*, John Wiley & Sons, Inc., 1997.
- Paolo Giudici, *Applied Data Mining: Statistical Methods for Business and Industry*, John Wiley & Sons Ltd., 2003.
- R. Johnson & D. Wichern, *Applied Multivariate Statistical Analysis*, 5th Edition, Prentice Hall, 2002.
- R. Roiger and M. Heatz, *Data Mining: A Tutorial Based Primer*, Addison Wesley, 2003.
- T&F Informa UK Ltd, “World Container Port Traffic League” , *Containerisation Yearbook 2006*, 2006.
- T. Kohonen, *Self-organization and Associative Memory*, 3rd edition, Springer-Verlag, Berlin, 1989.

부 록

〈표 1〉 주요 품목의 수입 물동량

(단위 : 천RT)

수입 물동량	유연탄	기타 광석	비철 금속	기계류	전기 기기	철강	화학 공업 생산물	플라스 틱고무	방직용 섬유	기타
1981	7,758	1,788	243	234	123	2,262	1,960	481	2,955	551
1982	9,484	2,251	294	284	149	866	2,156	432	3,103	1,307
1983	11,325	2,414	362	349	183	1,553	2,523	514	3,300	1,813
1984	13,563	2,697	461	445	233	2,733	2,936	581	3,429	1,853
1985	17,056	2,811	376	363	190	1,918	3,167	584	3,685	2,160
1986	17,881	2,607	369	356	186	2,804	3,963	800	4,377	2,038
1987	17,101	2,776	464	448	234	4,222	4,760	885	5,534	3,603
1988	21,516	2,735	441	425	223	5,085	6,604	970	5,365	5,428
1989	23,860	2,502	601	580	304	5,604	6,509	1,023	5,714	6,918
1990	24,592	2,631	722	696	364	6,790	7,053	982	5,774	11,446
1991	27,956	2,861	778	750	393	9,854	7,355	1,014	6,414	13,661
1992	28,497	3,041	889	858	449	6,683	6,973	1,047	6,361	7,918
1993	34,283	3,149	1,251	1,206	631	7,132	7,598	1,102	7,093	7,385
1994	36,726	8,878	2,335	2,252	1,179	11,604	8,132	1,188	7,361	4,275
1995	40,375	9,731	2,269	3,174	1,269	14,370	11,357	1,574	8,216	5,106
1996	44,697	9,611	1,835	2,970	1,585	15,490	11,313	1,423	9,563	6,321
1997	48,472	9,985	2,053	2,626	1,479	14,262	11,321	1,154	8,494	6,793
1998	50,355	9,500	1,498	2,133	824	5,955	9,535	820	5,853	4,705
1999	51,233	11,245	2,110	2,887	1,179	11,908	10,907	1,154	11,141	7,563
2000	60,487	11,738	1,994	2,724	1,319	15,068	10,224	1,199	15,027	8,786
2001	63,969	11,629	2,292	3,762	1,300	13,811	9,020	967	16,515	5,340
2002	63,777	14,058	2,271	5,636	1,635	17,534	9,316	1,506	20,285	5,782
2003	60,516	14,161	2,016	5,516	1,601	22,012	10,055	1,197	25,101	4,754
2004	67,841	15,044	2,829	6,098	1,726	26,441	10,560	1,055	31,030	4,734

〈표 2〉 주요 품목의 수출물동량

(단위 : 천RT)

수출물 동량	조제 식품	비철 금속	기계류	전기 기기	차량	철강	화학 공업 생산품	플라스 틱고무	방직용 섬유	기타
1981	102	47	200	272	312	5,193	351	725	1,236	8,911
1982	124	54	215	292	240	5,770	226	584	1,478	9,440
1983	111	52	212	288	300	6,144	466	615	1,767	9,670
1984	133	109	337	453	624	6,532	465	720	2,113	70,703
1985	175	182	496	663	1,476	6,675	565	910	2,527	70,237
1986	191	588	1,382	1,834	3,672	6,852	920	1,048	3,021	14,617
1987	390	886	2,032	2,692	6,552	7,968	855	1,133	3,613	18,521
1988	328	927	2,121	2,809	6,912	8,749	964	1,441	4,320	20,699
1989	373	362	888	1,181	4,272	8,886	903	1,422	5,166	19,787
1990	289	521	1,235	1,639	4,164	8,467	1,393	1,733	6,178	16,708
1991	291	326	809	1,076	4,680	8,421	1,883	2,384	7,387	14,535
1992	117	1,250	2,828	3,743	5,472	11,506	2,837	3,781	8,834	9,879
1993	37	2,035	4,540	6,004	7,668	12,473	3,309	4,467	10,563	5,742
1994	448	1,462	3,192	4,216	7,605	10,257	3,875	4,819	12,625	4,122
1995	463	1,427	4,054	4,322	10,752	10,948	5,048	4,512	14,866	5,363
1996	537	912	3,621	4,768	12,579	10,935	5,107	4,929	16,834	6,002
1997	695	1,570	4,027	5,168	13,949	11,947	6,998	4,836	18,953	7,184
1998	701	1,519	4,836	5,726	15,603	17,777	7,753	5,421	22,247	7,309
1999	533	1,326	4,654	5,449	17,343	14,472	8,969	5,344	25,115	9,106
2000	1,266	1,262	3,999	5,023	19,242	14,479	11,366	4,806	27,654	7,292
2001	1,526	1,500	5,389	5,981	18,365	14,565	10,716	5,287	28,697	6,022
2002	748	1,334	5,121	6,111	19,418	13,764	8,459	6,265	35,384	4,244
2003	955	1,681	5,262	5,735	21,701	15,080	8,999	4,406	44,735	4,906
2004	745	2,115	8,626	7,807	28,527	14,425	10,891	5,718	44,611	7,943

〈표 3〉 항만물동량 화물품목분류 신규 대비표

1994년 이전 분류		1994년 이후 분류		비고	
순번	17개 품목	순번	32개 품목	구성품목	HS코드
1	양곡 Grain	1	양곡 Cereals	날알곡물	제10류
2	유류 Oil	2	원유(역청유), 석유 Crude Petroleum Oils		2709
		3	석유정제품 Petroleum Gases and Other Gases	휘발유, 등유, 경유, 중유, 윤활유, 그리스 및 기타조제품	2710 2712 2713
		4	석유가스 및 기타가스류	LNG, LPG, 부탄, 에틸렌, 프로필렌, 부틸렌, 부타디엔, 석탄가스 등	2705 2711
3	유지류 Fats	5	동물식물성 유지류 Animal or Vegetable Fats and Oils		제15류
4	비료 Fertilizer	6	비료 Fertilizers		제31류
5	시멘트 Cement	7	시멘트 Cement	클링커, 포트랜드시멘트, 슬래그시멘트 등	2523
6	무연탄 Anthracite	8	무연탄 Anthracite		270111
7	유연탄 Bituminous Coal	9	유연탄 Bituminous Coal		270112
8	목재 Lumber	10	원목 Wood in The Rough		4403
		11	목재, 목탄, 코르크 등 Wood and Articles of Wood	목재, 목탄, 코르크, 짚 등의 농세공물 및 지조(枝條) 세공물	제44류 제45류 제46류 제47류
9	염류 Salt	12	기타광석 및 생산품 Other Ores	소금, 황, 토석류, 석고, 석회, 회 및 기타광석	제25류 제26류 제27류
10	인광석 Phosphate Rock				
11	기타광석 Other Ore	13	모래 Natural Sand		2505
12	수산물가공품 Marine Products	14	조제식품, 음료, 주류 등 Prepared Foodstuffs	육류·어류 등의 조제품, 코코아·곡물의 조제품, 조제식료품, 음료, 주류, 식료, 사료, 연초 등	제16류 제18류 제19류

					제20류 제21류 제22류 제23류 제24류
13	선어 Fresh Fishes	15	어패류, 갑각류 등 Fish, Crustaceans and Mollusks	어류, 갑각류, 연체동물 등	제3류
14	철광석 Iron Ore	16	철광석 Iron Ores		2601
15	기계류 Machinery	17	비철금속 및 그 제품 Base Metal and Articles Thereof	동, 니켈, 알루미늄, 연, 아연, 주석, 기타비금속 및 그 제품, 공구, 도구	제74류 제75류 제76류 제78류 제79류 제80류 제81류 제82류 제83류
		18	기계류 및 그 부품 Machinery and Mechanical Appliances	원자로, 보일러, 엔진, 크레인, 불도저, 에어컨, 냉장고, 인쇄기, 방적기, 직조기, 세탁기, 선반, 컴퓨터, 베어링 등, 광학기기, 사진용기기, 영화용기기, 측정, 검사기기, 의료용기기 및 이들의 부분품과 부속품	제84류 제90류
		19	전기기기 및 그 부품 Electrical Machinery and Parts Thereof	전동기, 발전기, 변압기, 축전지, 조명기기, 전기식 용접기기, 전기다리미, 전화기, 녹음기, 레이더기기, 무선기기, TV수상기, 축전기, 반도체, 케이블	제85류
		20	차량 및 그 부품 Vehicles and Parts Thereof	철도, 궤도용차량, 일반차량 및 부품, 교통신호기	제86류 제87류
		21	항공기선박 및 그 부품 Aircraft, Ships and Floating Structure	항공기·우주선 및 그 부품, 선박과 수상구조물	제88류 제89류
16	철재 Iron Material	22	고철 Scrap		7204 8908
		23	철강 및 그 제품 Iron Steel and Articles Thereof		제72류 제73류
17	기타 Others	24	육류 Meat		제2류

		25	제분공업생산물 Products of The Articles Thereof	밀가루, 전분	제11류
		26	기타 동·식물성 생산품 Other Animal and Vegetable Products	산동물, 낙동물, 산수목, 꽃, 채소, 과실, 견과류, 커피, 향신료, 채유용종자, 식물성엑스, 인산 등, 기타 동·식물성 생산품	제1류 제4류 제5류 제6류 제7류 제8류 제9류 제12류 제13류 제14류
		27	당류 Sugars	당류 및 설탕과자	제17류
		28	화학공업생산물 Products of The Chemical	무기·유기화합물, 의약품염료, 페인트, 잉크, 향료, 화장품, 비누, 계면활성제, 왁스, 카페인, 알부민, 전분 효소, 화약, 성냥, 필름, 인화지, 사진재료, 기타화학공업생산물	제28류 제29류 제30류 제32류 제33류 제34류 제35류 제36류 제37류 제38류
		29	플라스틱고무 및 그 제품 Plastics, Rubber and Articles Thereof		제39류 제40류
		30	피혁류 및 그 제품 Leather and Articles Thereof	원피, 가죽, 모피 및 그 제품	제41류 제42류 제43류
		31	방직용섬유 및 그 제품 Textiles and Textile Articles	견, 양모, 수모, 면, 기타식물성섬유, 워딩, 부직포, 인조스테이플섬유, 양탄자, 특수직물, 침구, 도포한직물, 편물의류, 기타섬유제품	제50류 제51류 제52류 제53류 제54류 제55류 제56류 제57류 제58류 제59류 제60류

					제61류 제62류 제63류
		32	기타 Other	펠프, 지류와 판지, 서적·서적, 신발, 모자, 우산, 지팡이, 조제우모, 돌·시멘트·석면제품, 도자기, 유리, 보석, 귀금속, 시계, 악기, 무기, 가구, 완구, 잡품, 폐기물	제48류 제49류 제64류 제65류 제66류 제67류 제68류 제69류 제70류 제71류 제91류 제92류 제93류 제94류 제95류 제96류 제97류

〈표 4〉 대미환율 실적치 및 전망

연도	원/달러(연평균)	연도	원/달러(연평균)
1981	681.3	2001	1,290.8
1982	731.5	2002	1,251.2
1983	776.2	2003	1,191.9
1984	806.0	2004	1,043.8
1985	870.5	2005	1,013.0
1986	881.3	2006	1,077.7
1987	822.4	2007	1,045.4
1988	730.5	2008	1,016.1
1989	671.4	2009	990.7
1990	708.0	2010	968.0
1991	733.6	2011	948.6
1992	780.8	2012	929.6
1993	802.7	2013	915.7
1994	803.6	2014	901.9
1995	771.0	2015	888.4
1996	804.8	2016	875.1
1997	951.1	2017	866.3
1998	1,398.9	2018	857.7
1999	1,189.5	2019	849.1
2000	1,130.6	2020	840.6

주 : 2005년까지의 자료는 실적치임.

자료 : 1. 한국은행(<http://www.bok.or.kr>) 통계시스템 자료.

2. 한국개발연구원, “중장기 경제성장 및 산업구조 전망”, 2004. 3.

〈표 5〉 생산 측면의 분야별 GDP(실질) 추이

(단위 : 십억 원, 2000년 불변가격)

구분	1970	1975	1980	1985	1990	1995	2000	2001	2002	2003	2004	2005
국내총생산	69,046	99,331	138,898	202,408	320,696	467,099	578,665	600,866	642,748	662,655	693,996	721,491
농업, 임업/ 어업	13,750	17,366	15,491	20,944	21,025	23,309	25,030	25,309	24,422	23,138	25,259	25,223
광업	1,955	2,861	3,250	3,491	3,227	2,332	2,037	2,035	1,879	1,895	1,947	1,917
제조업	5,019	12,011	22,562	38,419	69,525	103,279	151,243	154,503	166,243	175,417	194,886	208,519
전기,가스, 수도	247	560	1,303	2,758	5,505	8,884	13,213	14,169	15,258	15,981	17,035	18,390
건설업	4,695	6,806	12,566	18,504	33,769	47,925	42,927	45,279	46,529	50,549	51,459	51,522
도소매/ 숙박업	6,785	11,531	13,736	20,345	33,635	46,148	55,574	58,138	61,301	59,564	59,471	60,752
금융보험업	634	1,352	3,913	6,526	16,032	31,993	35,256	38,235	46,642	46,856	46,212	48,332
부동산/ 사업서비스	8,624	11,268	16,431	24,173	36,940	54,675	68,063	68,377	71,726	73,292	74,690	76,575
공공행정/ 국방	12,250	13,322	15,742	17,221	21,063	25,607	29,149	29,618	30,394	31,190	31,838	32,662
교육서비스	7,641	9,872	12,087	15,898	2,988	24,005	25,697	26,943	28,123	29,170	29,814	30,147
보건/ 사회복지	799	1,228	1,943	5,253	8,514	11,773	12,576	11,978	12,654	13,299	13,965	14,796
기타서비스	2,011	2,523	3,594	5,040	8,504	13,952	17,152	18,316	19,937	20,156	20,368	20,896
순생산물세 (조세-보조금)	2,633	5,052	9,412	14,179	27,998	50,730	64,611	66,441	72,312	74,663	76,244	78,868

자료 : 한국은행 경제통계시스템 (<http://ecos.bok.or.kr/>).

〈표 6〉 지출 측면의 분야별 GDP(실질) 추이

(단위 : 십억 원, 2000년 불변가격)

구분	1970	1975	1980	1985	1990	1995	2000	2001	2002	2003	2004	2005
국내총생산	69,046	99,331	138,898	202,408	320,696	467,099	578,665	600,866	642,748	662,655	693,996	721,491
최종소비지출	61,334	84,112	111,012	151,310	230,617	329,755	382,398	401,191	431,484	430,077	431,962	446,802
민간	47,214	65,589	86,397	121,937	185,863	270,628	312,301	327,685	353,560	349,200	348,067	359,273
가계	46,383	64,402	84,745	119,028	181,466	264,933	305,984	321,111	346,510	342,015	340,812	351,770
비영리단체	831	1,186	1,652	2,910	4,396	5,695	6,317	6,574	7,051	7,185	7,255	7,503
정부	14,120	18,523	24,615	29,372	44,754	59,127	70,098	73,507	77,924	80,877	83,895	87,529
총자본형성	10,998	19,057	35,988	54,417	115,458	179,701	179,413	179,334	189,898	194,579	203,859	206,632
총고정자본 형성	10,320	17,241	35,783	52,813	114,989	181,345	179,908	179,576	191,465	199,048	203,188	207,828
건설투자	8,020	12,219	22,228	33,945	71,619	105,066	96,676	102,438	107,883	116,428	117,729	118,161
설비투자	2,222	4,895	13,257	18,115	41,452	71,226	74,161	67,488	72,556	71,689	74,432	78,204
무형고정투자	78	127	298	754	1,918	5,053	9,071	9,650	11,025	10,931	11,027	11,462
재고증감	678	1,816	205	1,603	469	-1,644	-495	-243	-1,567	-4,469	671	-1,196
수출	3,076	9,326	19,915	32,184	55,670	112,547	236,210	229,764	260,221	300,824	359,710	390,418
(재화)	1,867	7,775	15,880	25,508	44,506	87,602	199,249	190,574	222,087	263,216	318,406	349,184
수입	6,856	13,325	27,351	36,124	77,742	153,631	217,979	208,899	240,665	264,930	301,719	322,531
(재화)	6,200	12,060	23,661	31,319	65,423	125,864	178,658	168,081	194,321	216,380	246,331	261,351
통계상불일치	494	162	-667	621	-3,306	-1,273	-1,377	-525	1,810	2,104	183	171

자료 : 한국은행 경제통계시스템 (<http://ecos.bok.or.kr/>).

〈표 7〉 주요 경제지표의 증가율 추이

(단위 : %, 2000년 불변가격 기준)

구분	1971	1975	1980	1985	1990	1995	2000	2001	2002	2003	2004	2005
국내총생산	8.2	5.9	-1.5	6.8	9.2	9.2	8.5	3.8	7	3.1	4.7	4
농림어업	3.9	4.3	-19.4	4.5	-6.5	5.3	1.2	1.1	-3.5	-5.3	9.2	-0.1
광공업	13.6	13.4	-1.3	6.4	8.3	11.2	16.7	2.1	7.4	5.5	11	6.9
(제조업)	17.8	13.4	-1.2	6.5	9.1	11.7	17	2.2	7.6	5.5	11.1	7
전기가스수도	21.6	15.2	23.2	11.1	17	6.1	12.8	7.2	7.7	4.7	6.6	8
건설업	-2.5	6.5	-3.6	4.3	25.5	7.7	-3.4	5.5	2.8	8.6	1.8	0.1
서비스업	9.1	4.6	3.4	7.9	7.8	8.1	6.1	4.8	7.8	1.6	1.9	3
순생산물세	20.9	4	-2.2	5.7	14.5	14.1	11.8	2.8	8.8	3.3	2.1	3.4
총고정자본형성	4.6	8.9	-10.7	5.3	25.4	13.1	12.2	-0.2	6.6	4	2.1	2.3
건설투자투자	-0.9	8	-4.5	4.5	30.5	9.5	-0.7	6	5.3	7.9	1.1	0.4
설비투자	24.4	11.3	-19.8	5.1	18.3	18.3	33.6	-9	7.5	-1.2	3.8	5.1
수출 (재화/서비스)	21.7	18.7	8.2	4.2	4.5	24.4	19.1	-2.7	13.3	15.6	19.6	8.5
수출(재화)	31.8	21.2	10.8	4.6	3.7	25.3	23	-4.4	16.5	18.5	21	9.7
수입 (재화/서비스)	19.1	2.1	-5	0.6	13.8	23	20.1	-4.2	15.2	10.1	13.9	6.9
수입(재화)	21.3	-0.6	-7.5	0.2	12.9	21.7	21.1	-5.9	15.6	11.4	13.8	6.1
GDP디플레이터	4.6	10.5	27.9	41.5	58.2	85.4	100	103.5	106.5	109.4	112.3	111.8

자료 : 한국은행 경제통계시스템 (<http://ecos.bok.or.kr/>).

〈표 8〉 국내총생산(GDP) 대비 수출(재화) 및 수입(재화)의 탄력성

(단위 : %, 2000년 불변가격 기준)

구분		1975	1980	1985	1990	1995	2000	2001	2002	2003	2004	2005
증가율	국내총생산	5.9	-1.5	6.8	9.2	9.2	8.5	3.8	7	3.1	4.7	4
	수출(재화)	21.2	10.8	4.6	3.7	25.3	23	-4.4	16.5	18.5	21	9.7
	수입(재화)	-0.6	-7.5	0.2	12.9	21.7	21.1	-5.9	15.6	11.4	13.8	6.1
탄력성	수출	3.59	7.20	0.68	0.40	2.75	2.71	-1.16	2.36	5.97	4.47	2.43
	수입	0.10	5.00	0.03	1.40	2.36	2.48	-1.55	2.23	3.68	2.94	1.53

〈표 9〉 대미달러 환율(평균자료) 변동추이

(단위 : 원/달러)

연도	1970	1975	1980	1985	1990	1995	2000	2001	2002	2003	2004	2005
환율	310.58	484.00	607.44	870.53	707.97	771.04	1130.61	1290.83	1251.24	1191.89	1144.67	1024.31

자료 : 한국은행 경제통계시스템 (<http://ecos.bok.or.kr/>).

〈표 10〉 인구 추이 및 전망

(단위 : 천 명)

추이	연도	1970년	1975년	1980년	1985년	1990년	1995년
	인구	32,241	35,281	38,124	40,806	42,869	45,093
	연도	2000년	2001년	2002년	2003년	2004년	2005년
	인구	47,008	47,354	47,615	47,849	48,082	48,294
전망	연도	2006년	2008년	2011년	2013년	2015년	2020년
	인구	48,497	48,877	49,375	49,626	49,802	49,956

자료 : KOSIS(통계청) 통계정보시스템 (<http://kosis.nso.go.kr/>).

데이터 마이닝 기법을 이용한 항만물동량 예측 활용방안 연구

2006年 12月 26日 印刷

2006年 12月 28日 發行

編輯兼
發行人 李 正 煥

發行處 韓國 海 洋 水 產 開 發 院
서울특별시 서초구 방배3동 1027-4
수암빌딩

전 화 2105-2700 FAX : 2105-2800

등 록 1984년 8월 6일 제16-80호

組版・印刷 / 해항사 393-0836

정가 15,000원

판매 및 보급 : 정부간행물판매센터

Tel : 394-0337, 734-6818